CHARACTERIZATION OF GENES INVOLVED IN THE METABOLISM OF ATHEROGENIC LIPOPROTEINS IN THE MOUSE

Mariëtte Hoffer

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Hoffer, Maria Johanna Veronica

Characterization of genes involved in the metabolism of atherogenic lipoproteins in the mouse / Maria Johanna Veronica Hoffer. -[S.l. : s.n.]. - Ill.
Proefschrift Leiden. - Met lit. opg. Met een samenvatting in het Nederlands.
ISBN 90-5412-013-4
Trefw.: lipoproteïne metabolisme / moleculaire biologie.

STELLINGEN

1 Bij het onderzoek naar risicofactoren voor hart- en vaatziekten bieden genetisch gemodificeerde muize modellen een extra mogelijkheid om een "brug te slaan" tussen fibrinolyse en stolling enerzijds en het lipoproteïne metabolisme anderzijds.

2 Kruising van de *Ldlr* knockout muis met een *Apoe* knockout muis zal een muis model opleveren dat van nut kan zijn bij de karakterisatie van de chylomicron remnant receptor.

3 De deeltjes in de IDL/LDL fracties bij scheiding van muize serum met behulp van superose-6 kolomchromatografie zijn wel qua grootte maar niet qua samenstelling identiek aan de gelijknamige humane deeltjes. (Yokode et al., Science 1990; 1250: 1273-1275 en Ishibashi et al., J. Clin. Invest. 1993; 92: 883-893).

4 De conclusie van Jäckle et al. betreffende de rol van de 'LDL receptor-gerelateerde eiwit' (LRP) bij de klaring van triglyceride-rijke lipoproteïnen in de rat gaat voorbij aan de reeds eerder gepubliceerde resultaten van van Dijk et al.. (Jäckle et al., J. Lipid Res. 1993; 34:309-314 en van Dijk et al., Biochem. J. 1991; 279:863-870).

5 De door Shen en Howlett veronderstelde aminozuur samenstelling van het ECL eiwit in de rat wordt onvoldoende ondersteund door hun experimentele resultaten. (Shen en Howlett, Arch. Biochem. Biophys. 1992; 297:345-353)

6 Door de ziekte van Alzheimer in verband te brengen met het apolipoproteïne E polymorfisme, lijkt de "blood-brain barrier" doorbroken. (Strittmatter et al., Proc. Natl. Acad. Sci, USA 1993; 90:1977-1981).

7 Zonder een betrouwbare genetische diagnostiek is de herkenning van aangedane familieleden van een FCH patiënt een welhaast onmogelijke zaak.

8 Door het reproduceren van de hoofdstukken 2, 3, 4, en 6 in de reeds gepubliceerde vorm zijn voor de totale oplage van dit proefschrift 13000 pagina's bespaard.

9 Het oude spreekwoord "*Wat een boer niet kent dat eet hij niet*" zou tegenwoordig beter kunnen luiden " *Wat een boer goed kent dat eet hij niet*".

10 In zijn huidige vorm lijkt de "*Annex*" meer een "ex-" dan een "toekomstig" laboratorium.

11 Wie het laatst lacht is langzaam van begrip.

Stellingen bij het proefschrift: Characterization of genes involved in the metabolism of atherogenic lipoproteins in the mouse.

Alphen aan den Rijn, 20 september 1993 Mariëtte Hoffer

CHARACTERIZATION OF GENES INVOLVED IN THE METABOLISM OF ATHEROGENIC LIPOPROTEINS IN THE MOUSE

Proefschrift

ter verkrijging van de graad van Doctor
aan de Rijksuniversiteit te Leiden,
op gezag van de Rector Magnificus Dr. L. Leertouwer,
hoogleraar in de faculteit der Godgeleerdheid,
volgens besluit van het College van Dekanen
te verdedigen op donderdag 25 november 1993
te klokke 14.15 uur

door

Maria Johanna Veronica Hoffer

geboren te Wieringermeer in 1965

PROMOTIECOMMISSIE

Promotor:	Prof. Dr. G.J.B. van Ommen
Co-promotoren:	Dr. M.H. Hofker
	Dr. L.M. Havekes (IVVO-TNO, Gaubius Laboratorium, Leiden)
Referent:	Prof. Dr. B. Wieringa (Katholieke Universiteit Nijmegen)
Overige leden:	Prof. Dr. Th.J.C. van Berkel
	Prof. Dr. P. Brakman
	Prof. Dr. Ir. P.H.M. Lohman

The studies presented in this thesis were performed at the Department of Human Genetics of the Leiden University, Leiden, in collaboration with IVVO-TNO, Gaubius Laboratory, Leiden, and was supported by HGO-TNO (# 900-539-052).
Financial support by the Netherlands Heart Foundation for the publication of this thesis is gratefully acknowledged. Additional support was obtained from IVVO-TNO, Gaubius Laboratory, Leiden.
Printed by Mosterd & Van Onderen!

Voor Peter

CHAPTER 1

GENERAL INTRODUCTION

1. INTRODUCTION

Lipoproteins are particles consisting of lipid and proteins, which play an important role in the transport of cholesterol and triglycerides through the body. Variation in the levels and structure of circulating lipoproteins are due to heterogeneity in the environmental and genetic background in man. Mutations in genes involved in the lipoprotein metabolism may influence the processing of lipoprotein particles and their plasma concentrations. Interestingly, the majority of single gene defects display a variable influence on lipoprotein levels due to additional genetic factors. Since variations in lipoprotein levels are strongly associated with the susceptibility to atherosclerosis, the characterization of these hereditary influences will contribute in our knowledge about the genetic components involved in this major human disease.

In contrast to human studies, studies with animal models can be done under controlled environmental conditions and against a homogenous genetic background. For genetic studies the mouse is the animal model of choice. Mouse inbred strains are a source of genetically identical individuals and have shown to be important for the identification of genes controlling the lipoprotein structure. To study the influences of genetic factors in more detail characterization of genes known to be involved in the human lipoprotein metabolism is required. A good interpretation of the results found with animal models strongly depends on our knowledge about the mouse lipoprotein metabolism and its associated genes.

2. THE LIPOPROTEIN METABOLISM

Cholesterol is essential for cell membranes and is used for synthesis of steroid hormones and bile acids, whereas triglycerides are used as a source of energy for muscles and storage in adipose tissues. Cholesterol and triglycerides are transported in the blood within lipoproteins because they are insoluble in an aqueous environment. Lipoproteins are particles consisting of lipids (cholesterol, cholesteryl esters, triglycerides and phospholipids) and proteins, called apolipoproteins. The surface of lipoproteins contains phospholipids, free cholesterol and apolipoproteins while the core of these particles consists mainly of cholesteryl esters and triglycerides. The major lipoprotein classes, chylomicrons, very low density lipoproteins (VLDL), low density lipoproteins (LDL), and high density lipoproteins (HDL), are heterogenous particles

Table 1. Biophysical and chemical properties of human plasma lipoproteins[#].

	Chylomicrons	VLDL	LDL	HDL
Density (g/ml)	< 0.96	0.96-1.006	1.019-1.063	1.063-1.210
Electrophoretic mobility	origin	pre-β	β	α
Triglycerides (% wt)	88	56	13	15
Phospholipids (% wt)	8	20	28	45
Cholesteryl ester (% wt)	3	15	48	30
Free cholesterol (% wt)	1	8	10	10
Protein (% wt)	1-2	6-10	21	45-55
Major apolipoproteins	A1, A4, B48, C1, C3, E	B100, C1, C2, C3, E	B100	A1, A2, E

[#]Gotto et al., 1986

which differ in density, size and electrical charge, and vary in their apolipoprotein composition (Table 1). This heterogeneity leads to differences in their metabolism and mode of cell interaction.

The lipoprotein metabolism can be divided into three different pathways: i) the exogenous lipid transport (Fig. 1A), ii) the endogenous lipid transport (Fig. 1B), and iii) the reverse cholesterol transport (Fig. 1C), (for review, see Mahley et al., 1984; Gotto et al., 1986; Breslow, 1988; Havel and Kane, 1989; Eisenberg, 1990).

2.1 The exogenous lipid transport

Dietary cholesterol and triglycerides are absorbed and packed in the intestinal mucosal cells into large particles called chylomicrons. These very large triglyceride-rich lipoproteins, containing predominantly apolipoprotein (apo) B48, apoA1 and apoA4, are secreted into the lymph. After entering the circulation, chylomicrons acquire apoE, apoC1, apoC2 and apoC3 from HDL. These modified chylomicrons are rapidly hydrolysed by endothelial lipoprotein lipase (LPL) with apoC2 serving as a cofactor. During this lipolysis process, excess surface components, consisting of phospholipids and apolipoproteins, are transferred to HDL. The remaining lipoprotein particles are called chylomicron remnants. These particles are enriched in cholesterol

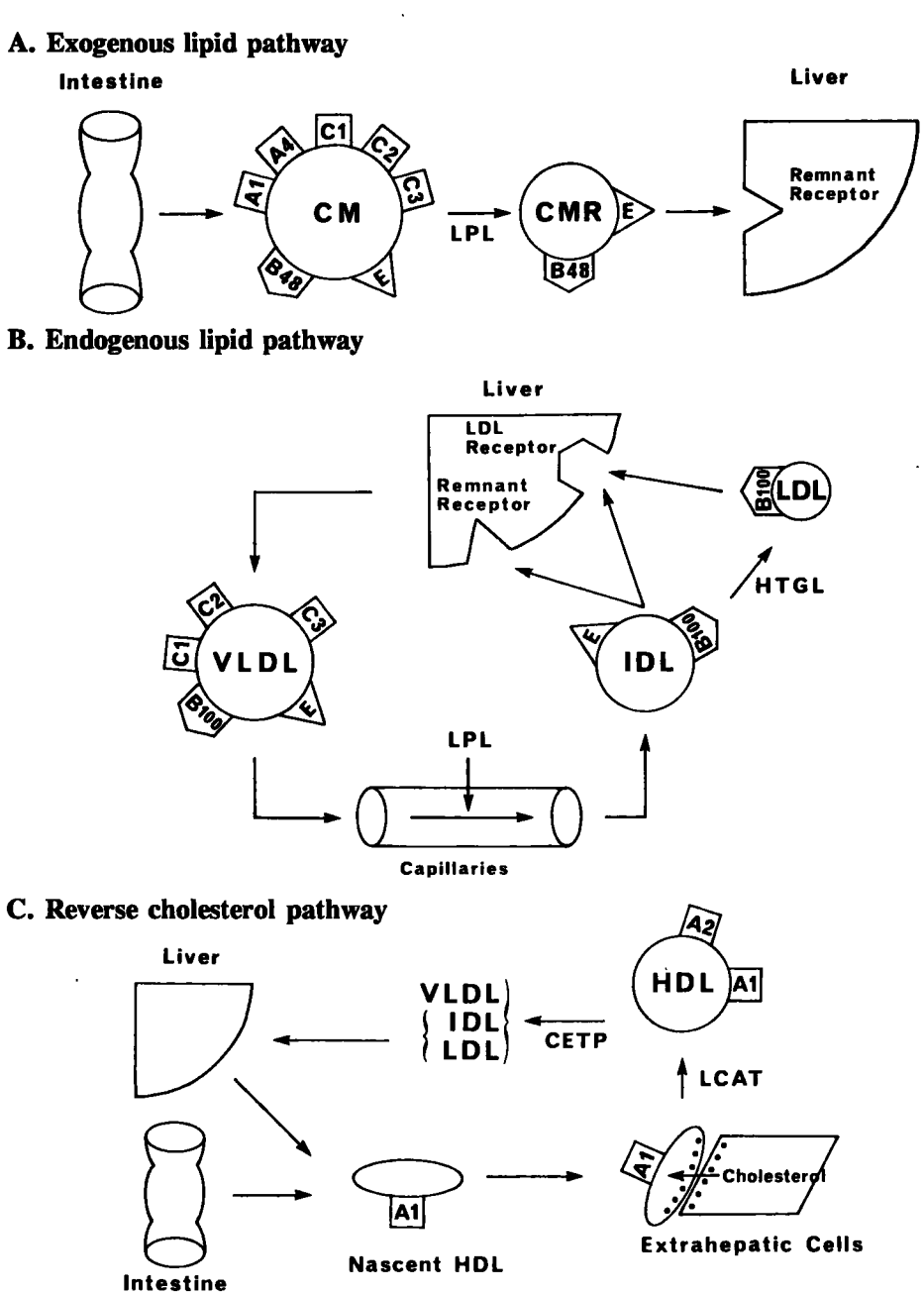

Fig. 1 Schematic representation of the lipid transport pathways. CM: chylomicron; CMR: chylomicron remnant. Redrawn from Breslow (1988).

and contain mainly apoB48 and apoE. Finally, the chylomicron remnants are taken up by the liver through receptor-mediated clearance. ApoE serves as a ligand for binding of these remnants to the putative apoE or chylomicron-remnant receptor.
Dietary cholesterol and triglycerides taken up by the liver through the exogenous pathway are used for the synthesis of bile acids and cholesterol.

2.2 The endogenous lipid transport

Endogenous synthesis of triglycerides and cholesterol occurs primarily in the liver, where these lipids are packed and secreted as nascent VLDL. VLDL particles are, like chylomicrons, rich in triglycerides. VLDL contains newly synthesized apoB100 and small amounts of apoE, apoC1, apoC2, and apoC3. Similar to chylomicrons, triglycerides are hydrolysed by LPL resulting in the formation of VLDL remnants or intermediate density lipoproteins (IDL). IDL can be cleared from the circulation by the liver using both the putative apoE receptor and the LDL receptor which recognize and bind apoE. IDL which escapes rapid receptor-mediated clearance by the liver will be further processed by lipolysis involving hepatic triglyceride lipase (HTGL) which leads to the formation of LDL. LDL is a cholesterol-rich particle carrying apoB100 as the only protein component on its surface. ApoB100 is the ligand for binding to the LDL receptor. Most of the LDL will be cleared from the circulation by the liver.

2.3 The reverse cholesterol transport

The third lipoprotein pathway, in which HDL plays a major role, is called the reverse cholesterol transport. In this process nascent HDL, phospholipid disks containing apoA1, is synthesized and secreted by intestine and liver. ApoA1 is a cofactor for lecithin:cholesterol acyltransferase (LCAT). In plasma, free cholesterol from extra-hepatic cells is esterified by LCAT. The produced cholesteryl esters are transported to the core of HDL thereby changing its shape into a spherical HDL (HDL_3). During lipolysis of triglyceride-rich particles, apolipoproteins, mainly apoA2, and phospholipids are transferred into HDL. These small HDL_3 particles increase in size by additional action of LCAT, forming HDL_2. Cholesteryl esters in HDL_2 are transferred to apoB-containing lipoproteins, VLDL, IDL and LDL, by cholesterol ester transfer protein (CETP). The removal of cholesteryl esters from HDL_2 is accompanied by a reciprocal transfer of triglycerides by CETP from triglyceride-rich lipoproteins to HDL. The triglyceride-rich HDL is remodelled by the action of HTGL, which hydrolysis triglycerides and phospholipids of HDL. This leads to the formation of

smaller HDL particles (HDL_3) and the dissociation of some apolipoproteins from HDL (Tall, 1992). The redistribution of cholesteryl esters from peripheral tissues to apoB-containing lipoproteins allows their uptake by the liver through LDL receptor-mediated clearance. This pathway, the reverse transport of cholesterol, is involved in the uptake of cholesteryl esters via the LDL receptor pathway in the liver and subsequent removal of excess of cholesterol, partly as bile acid, from the body via the intestine.

3. GENETIC ASPECTS OF THE LIPOPROTEIN METABOLISM

From epidemiological studies it has been estimated that 50% of the variation in plasma levels of lipoproteins in man has a genetic background. Among the genetic factors known to influence the concentration and metabolism of lipoprotein particles are: i) specific receptors involved in binding and internalisation of lipoproteins, ii) apolipoproteins, and iii) enzymes involved in the modification of lipoproteins.

Disturbances in the lipoprotein metabolism leading to an accumulation of specific lipoproteins in the plasma, are often the result of mutations in these genes (for review, see Havel and Kane, 1989). The most atherogenic lipoproteins are LDL and β-VLDL. β-VLDL is derived from chylomicron- or VLDL-remnants. When removal of these lipoproteins is hampered, these particles accumulate in the plasma and become enriched in cholesterol. Elevated plasma levels of LDL and β-VLDL form a major risk factor for atherosclerosis, which is characterized by foam cell formation in the intima of the arterial wall.

This thesis will focus on the APOE-C1-C2 gene cluster and the LDLR gene which play an important role in the metabolism of atherogenic lipoproteins. Genetic disorders affecting the expression or function of apoE and the LDL receptor are familial dysbetalipoproteinemia (FD) and familial hypercholesterolemia (FH), respectively. FD is associated with mutations in apoE which lead to an elevated chylomicron- and VLDL-remnant concentration in the plasma (Mahley and Rall, 1989). FH leads to an accumulation of plasma LDL cholesterol while the concentration of other plasma lipoproteins is not affected (Goldstein and Brown, 1989). Mutations in the APOC2 gene lead to hypertriglyceridemia which is characterized by an accumulation of triglyceride-rich lipoproteins. Sofar no primary qualitative or quantitative abnormalities of human apoC1 have been reported.

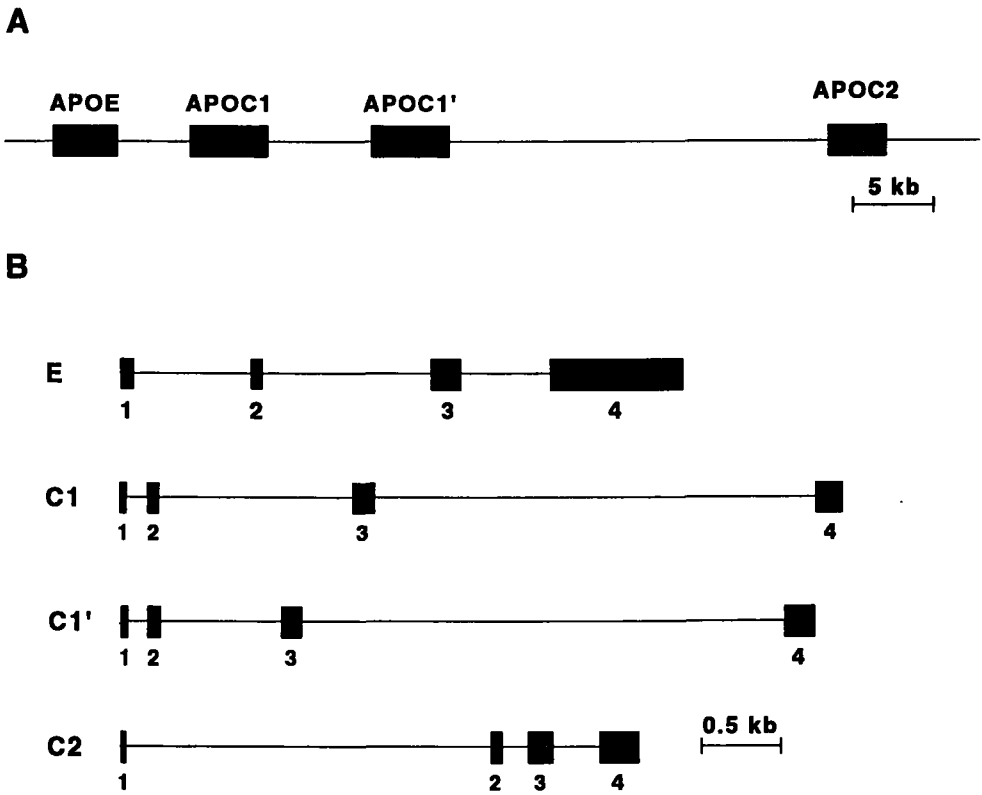

Fig. 2 Structure of the apolipoprotein E-C1-C2 gene cluster and individual genes. (A) The APOE-C1-C2 gene cluster. Filled boxes indicate genes. (B) The exon-intron structures of the apolipoprotein genes; APOE, APOC1, APOC1' and APOC2.

3.1 The APOE-C1-C2 gene cluster

APOE, APOC1 and APOC2 are three of the seven members of the apolipoprotein gene family, which also includes APOA1, APOA2, APOA4, and APOC3 (Li et al., 1988). The proteins expressed by these genes are characterized by extensive regions of amphipatic helix. In humans, APOE is contained together with APOC1, pseudo APOC1, and APOC2 within a 48 kb gene cluster on the long arm of chromosome 19, 19q12-19q13.2 (Fig. 2A) (Brook et al., 1987; Myklebost and Rogne, 1988; Smit et al., 1988a).

The APOE gene; structure and expression

The human APOE gene has been characterized extensively. The gene spans a region of approximately 3.6 kb and consists of four exons (Fig. 2B)(Das et al., 1985; Paik et al., 1985). The mRNA encodes a protein of 317 amino acids including a signal peptide of 18 amino acid residues (Rall et al., 1982; Wallis et al., 1983; McLean et al., 1984). ApoE is a single polypeptide with a molecular weight of 34 kD. There are 3 major isoforms found encoding apoE2, E3 and E4 (Zannis et al., 1981, 1982; Utermann et al., 1982). These isoforms are the result of amino acid changes leading to charge differences. Population studies have shown that this allelic variation causes small but significant changes in plasma LDL levels (Smit et al., 1988b). Most of the FD patients have the apoE2E2 phenotype but only a very small percentage of E2E2 individuals will develop FD. This indicates that additional factors are required for the development of FD (Utermann, 1985).

ApoE is synthesized primarily by the liver but it is also produced by a variety of non-hepatic tissues including brain, kidney and adrenal glands (Wallis et al., 1983; Blue et al., 1983). The tissue specific synthesis of apoE is controlled by an array of elements found in the immediate 5'-flanking region of the APOE gene and throughout the whole APOE-C1-C2 gene cluster (For review, see Taylor et al., 1991; Zannis et al., 1992).

Function of apoE

ApoE which is present on chylomicron- and VLDL-remnants, plays an important role as a ligand for binding of the particles to the LDL receptor and the putative apoE receptor. The protein is organized as two distinct structural domains, each with different functions. The amino-terminal part contains the region of apoE that binds to the LDL receptor and the putative apoE receptor. The three dimensional structure has been determined by crystallography, showing that this domain consists of four helices (Wilson et al., 1991).

The carboxy-terminal part of apoE represents the major lipid-binding region which is composed of a long stretch of α-helices (Weisgraber et al., 1986). In addition, this carboxy-terminal part also contains a heparin binding domain which is thought to play a role in the hydrolysis of triglyceride-rich lipoproteins by anchoring the particle to endothelial heparan-proteoglycans.

The APOC1 gene; structure and expression

The human APOC1 gene is located 4.3 kb downstream of the APOE gene (Myklebost and Rogne, 1986). One copy of the APOC1 gene, pseudo APOC1, is located 7.5 kb downstream of APOC1 (Davison et al., 1986; Lauer et al., 1988). The APOC1 gene spans a region of approximately 4.6 kb and is composed of four exons (Fig. 2B). Nucleotide sequence analysis of cloned cDNA (Knott et al., 1984) indicated that apoC1 is synthesized with a 26-residue signal peptide that is removed during intracellular processing. Plasma apoC1 consists of 57 amino acids residues in a single polypeptide chain with a calculated molecular weight of 6.6 kD (Jackson et al., 1974; Shulman et al., 1975). The APOC1 gene is expressed primarily in the liver but expression has also been detected when monocytes differentiate into macrophages (Lauer et al., 1988).

No mRNA products of the pseudo APOC1 gene can be detected in any tissue. This is probably the result of a point mutation in exon 3 which introduces a translation stop signal (Lauer et al., 1988). Although the coding region is small (249 bp), human APOC1 is 4.6 kb in length due to large introns that contain a total of 9 *Alu* repeats. Basepair changes in these *Alu* repeats were used to estimate that the divergence between the APOC1 gene and the pseudo APOC1 gene took place after the divergence of rodents and primates which occurred 40 million years ago (Raisonnier, 1991).

Function of apoC1

ApoC1 is a constituent of chylomicrons, VLDL, and HDL (Breslow, 1988) and it has been shown to activate LCAT *in vitro* but not as efficiently as apoA1 (Soutar et al., 1975). Using synthetic peptides of apoC1, studies indicated that a fragment consisting of residue 17 to 57, contains all of the structural features necessary to activate LCAT and is as active as the intact protein. Residues 32-57 represent one of the major phospholipid-binding regions of apoC1 (Soutar et al., 1978). Another function of apoC1 is that it inhibits the action of phospholipase A (Poensgen, 1990).

In addition, studies showed that apoE and apoC1 have an opposite effect on remnant uptake by the liver (Windler and Havel, 1985). Studies with rat liver membranes showed that the apoE-dependent binding of β-VLDL to the LDL-receptor related protein is blocked by apoC1 (Kowal et al., 1990; Weisgraber et al., 1990). Also studies with the LDL receptor showed that apoC1 inhibits effectively the binding of apoE to this receptor (Sehayek and Eisenberg, 1991).

The APOC2 gene; structure and expression

The human APOC2 gene has been isolated and sequenced (Wei et al., 1985; Das et al., 1987). The gene consists of four exons and three introns and spans a region of approximately 3.4 kb (Fig. 2B). Intron 1 contains a polymorphic CA-repeat, which varies in length between different APOC2 alleles. Intron 3 is composed largely of 6 nearly perfect copies of a 37-bp tandem repeat. The same core sequence is found at other loci in the genome but all appear to be located at the long arm of chromosome 19 (Das et al., 1987). The amino acid sequence has been reported (Jackson et al., 1977; Hospattankar et al., 1984) and confirmed by nucleotide sequencing of cDNA clones (Myklebost et al., 1984; Jackson et al., 1984; Fojo et al., 1984; Sharpe et al., 1984). The protein has a 22-residue signal peptide that is cotranslationally cleaved (Sharpe et al., 1984). ApoC2 is a single polypeptide chain consisting of 79 amino acids with a calculated molecular weight of 8.8 kD. The APOC2 gene is expressed primarily in liver and intestine (Myklebost et al., 1984).

Function of apoC2

ApoC2 is a constituent of chylomicrons, VLDL and HDL. ApoC2 is the cofactor for the enzyme LPL which catalyses the hydrolysis of triglycerides in chylomicrons and VLDL (LaRosa et al., 1970). The physiological importance of apoC2 in activating LPL has been established by the finding of homozygous patients with inherited apoC2 deficiency who have severe hypertriglyceridemia (Breckenridge et al., 1978).

The structure of apoC2 is predicted to contain three helical regions between residue 13 to 22, 29 to 40, and 43 to 52 which are thought to be involved in phospholipid binding (Catapano et al., 1979). Studies using synthetic peptides of apoC2 have shown that LPL interacts with the COOH-terminal amino acids 56-79 (Kinunnen et al., 1977). Deletion of the COOH-terminal tetra peptide, residues 76-79, prevents the protein from activating LPL and is probably important in mediating the initial interaction between apoC2 and LPL (Cheng et al., 1990).

3.2 The LDLR gene

The LDLR gene; structure and expression

The LDLR gene consists of 18 exons spanning a region of 45 kb on the short arm of chromosome 19 (19p13.3) (Südhof et al., 1985; Lindgren et al., 1985). There is a close correlation between functional domains in the LDL receptor protein and the

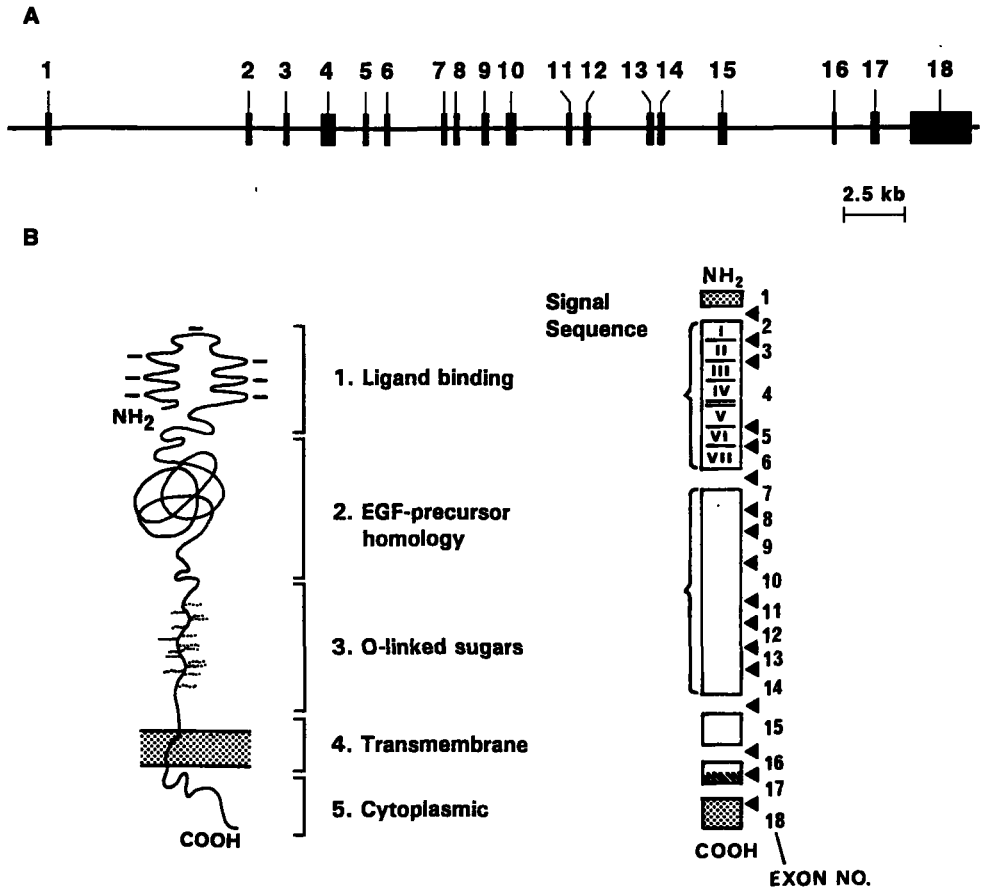

Fig. 3 Schematic representation of the LDLR gene and protein. (A) The exon-intron structure of the human LDLR gene. (B) The LDL receptor structure: correspondence between the structures of the protein and the gene. Closed triangles indicate introns. (Redrawn from Südhof et al., 1985 and Brown and Goldstein, 1986).

exon-intron organization of the gene (Fig. 3) (Südhof et al., 1985). The mRNA has a size of 5.3 kb, but only 2.8 kb encode an 839 amino acids protein with an additional signal peptide of 21-amino acid residues. The remaining 2.5 kb of the mRNA corresponds to the 3'-untranslated region which includes three copies of an *Alu* repeat (Yamamato et al., 1984). The human LDL receptor is a transmembrane glycoprotein that is synthesized as a precursor with a molecular weight of 120 kD (Tollenshaug et al., 1982). During posttranslational modification in the Golgi apparatus, the molecular weight increases from 120 to 160 kD due to glycosylation (Schneider et al., 1982).

LDL receptors are present on the surface of essential all cultured mammalian cells. In vivo, however, the most important site of expression is the liver (Brown and Goldstein, 1986). The promoter region of the LDL receptor has been characterized and several important cis-acting regulatory elements have been detected. There are three direct imperfect repeats, each 16 bp in length, and a TATA-like sequence. Each of these elements is essential for full promoter activity. Repeats 1 and 3 contain high-affinity binding sites for the positive transcription factor Sp1. Repeat 2 does not bind Sp1 and contains the sterol regulatory element (SRE1), that is required for sterol-mediated repression. Most studies point to a complicated mechanism for promoter activity and sterol-mediated regulation (Südhof et al., 1987a, 1987b; Dawson et al., 1988; Smith et al., 1990b). Additional factors such as cellular growth and hormonal control may regulate LDL receptor transcription but the regulatory elements involved in this regulation are still unknown (Auwerx et al., 1989; Wade et al., 1989; Mazzone et al., 1990).

Function of the LDL receptor.

The LDL receptor protein consists of five structural domains. From the NH_2-terminus to the COOH-terminus the following domains can be discerned: (i) the ligand binding domain, (ii) the EGF precursor domain, (iii) the O-linked sugar domain, (iv) the transmembrane spanning domain, and (v) the cytoplasmic domain (Yamamoto et al., 1984). The function of these domains have been characterized by studies on numerous naturally occurring mutants and by site directed mutagenesis (For reviews see Hobbs et al., 1990, 1992; Soutar, 1992).

(i) The ligand binding domain consists of 292 amino acids encoded by exons 2-6 and comprises seven copies of a negatively charged cysteine-rich repeat. This domain is involved in the binding of lipoproteins containing apoB100 and apoE (Innerarity et al., 1984; Yamamoto et al., 1984). Due to the conformation of this domain the negatively charged residues interact with the positively charged clusters on apoE or apoB100. Studies on this domain, using site directed mutagenesis, showed that repeats 3-7 are involved in the binding of LDL while repeat 5 is involved in the binding of both LDL and β-VLDL (Esser et al., 1988; Russell et al., 1989).

(ii) The EGF precursor domain contains approximately 400 amino acids and is encoded by exons 7-14. Three repeated sequences (A, B and C) are found in this region and show homology with repeat sequences in the EGF precursor of the mouse (Russell et al., 1984; Yamamoto et al., 1984; Südhof et al., 1985). This domain is

required for the release of bound ligands from the receptor in the endosomal compartments, which is essential for recycling of the LDL receptor. Repeat C is essential for binding of LDL and it is suggested that it acts as a spacer region that extends the receptor-binding domain away from the cell surface (Davis et al., 1987).

(iii) The O-linked sugar domain is encoded by exon 15 and comprises a stretch of 58 amino acids, enriched in serine and threonine residues of which the majority is glycosylated. Proteolytic studies have revealed that this region contains the clustered O-linked sugar chains. The function of this domain remains unclear (Davis et al., 1986; Russell et al., 1984).

(iv) The transmembrane spanning domain is encoded by exon 16 and a part of 17 and consists of 22 hydrophobic amino acids. This domain is believed to anchor the receptor in the cell surface by spanning the plasma membrane (Russell et al., 1984).

(v) The last domain of the LDL receptor is the COOH-terminal cytoplasmic domain of 50 amino acid residues and is encoded by parts of exons 17 and 18. This domain plays an important role in clustering in coated pits on the cell surface thereby facilitating endocytosis. A tetrameric sequence Asn-Pro-x-Tyr (NPxY, x stands for any amino acid) in this domain is found to be important for the coated pit internalization. This NPxY sequence has been found in several other surface receptors. It is suggested that the NPxY sequence has a conditional role in ligand-independent internalization of these receptors (Chen et al., 1990).

4. THE MOUSE AS AN ANIMAL MODEL FOR STUDYING THE LIPOPROTEIN METABOLISM

The variability in plasma levels of lipids and lipoproteins is due to heterogeneity in environmental and genetic background. In man, about 50% of this variability is due to multiple genetic factors. Even single gene defects display a variable influence on lipid and apolipoprotein levels because of confounding interactions with environmental and/or genetic factors. This complexity of interacting factors influencing the lipoprotein metabolism makes it difficult to identify these factors individually. As a consequence, studies of plasma lipoproteins or atherosclerosis in man are hampered. However, the use of animal models could circumvent some of these limitations of lipoprotein studies in man by providing a defined genetic background and controlled environmental conditions.

4.1 Advantages of a mouse model

As in most areas of human biology, studies on human plasma lipoproteins or atherosclerosis have been enriched and completed by investigating animal models. The use of animal models will increase our knowledge about the interactions between different genetic factors and between genetic and environmental factors involved in the lipoprotein metabolism. However, one has to consider that there are differences between species with respect to quantitative traits, metabolic pathways, and developmental pathways. Animal models do not serve to duplicate precisely genetic variations that occur in humans but rather to gain a better understanding of the biochemical mechanisms and the types of polymorphisms that underlie heritable variations of lipoproteins. In addition, animal models also provide a means for identifying new genes affecting lipid metabolism and for defining in detail the molecular mechanisms involved in lipid transport. Generally used species which serve as models for the study of plasma lipoproteins as well as atherosclerosis include rabbits, hamsters, mice, rats, pigs and non-human primates. (For recent reviews see Armstrong and Heistad, 1990; Reue et al., 1990; Overturf and Loose-Mitchell, 1992).

The selection of a model strongly depends upon the subject of investigation. Animal models with spontaneous mutations affecting various aspects of lipoprotein metabolism have been valuable in the study of genetic and biochemical basis of similar monogenic disorders in humans. Examples are hypercholesterolemia in the Watanabe heritable hyperlipidemic (WHHL) rabbit with a defect in the LDLR gene (Yamamoto et al., 1986) and atherosclerosis in pigs with genetic variation in their apolipoproteins (Rapacz et al., 1986). A problem associated with these animal models is that they are very rare. In addition, some animal species are in favour because their lipoprotein metabolism closely resembles that of humans. These animals are used for studying lipoproteins and the influences of environmental factors such as diet, on the composition of these lipoproteins. A disadvantage of these models is that they mainly represent large animals, like monkeys. These models are expensive to maintain and genetic studies are difficult due to a long generation time and, as in man, heterogeneity in genetic background.

For genetic studies the mammal of choice is the mouse since it offers a variety of advantages when compared with other animal models. There are hundreds of different inbred strains available, each strain representing a unique gene pool in which natural polymorphisms have been fixed by inbreeding. In addition, several sets of recombinant inbred strains have become available as well. Inbred strains of mice form a source of

infinite numbers of genetically identical individuals. Therefore, the mouse provides an animal model system for examining the heritable variation of lipoproteins since genetic and biochemical analyses are greatly simplified and environmental influences can be controlled. Another major advantage is that the genetic map of the mouse is well defined however, only a small number of genes has been characterized in more detail (Lusis and Sparks, 1989; Hillyard et al., 1993). More recently, special techniques have become available for manipulating the genes of interest by transgenesis and gene targeting. The technology is now available to generate animal models with specific alterations in the genes of interest, independent of spontaneous variations (see below).

For a long time, mice were not used commonly to study lipoprotein metabolism and atherosclerosis, because early attempts to produce atherosclerosis were discouraging due to variable expression of the disease and mortality caused by the experimental diet. The identification of inbred strains of mice, either genetically susceptible or resistant to diet-induced atherosclerosis, resolved some of the problems associated with the use of mice as models for atherosclerosis (Roberts and Thompson, 1976; Morrisett et al., 1982; Paigen et al., 1985). The development of appropriate experimental diets have resulted in reproducible atherosclerotic lesions in the mouse (Nishina et al., 1990; Steward-Phillips and Lough, 1991). Due to these characteristics the mouse is widely used as an animal model for studying the interaction between genes and environmental factors on the metabolism of lipoproteins and atherosclerosis.

4.2 Genetic aspects of the mouse lipoprotein metabolism

The recent development of highly polymorphic genetic markers, simple sequence repeats, resulted in a large set of genetic markers spanning randomly the whole mouse genome (Cornall et al., 1991). The genetic map of the mouse is well defined and it is comprised of a large number of loci/genes including genes involved in the lipoprotein metabolism (Table 2)(Hillyard et al., 1993). Comparison of the linkage maps in man and mouse shows that the loci for the genes involved in the lipoprotein metabolism are contained within homologous linkage groups in mice and man, suggesting a common evolutionary pathway. Mouse and man have essentially the same set of genes encoding enzymes and receptors directing the major lipid transport pathways. The mouse lipoproteins and apolipoproteins appeared to be structurally and functionally similar to the human counterparts (LeBoef et al., 1983; Camus et al., 1983). The homology between human and mouse gene maps will contribute to the mapping of new genes affecting the lipoprotein metabolism in humans. New loci mapped in the mouse could

Table 2. Chromosomal localization of genes involved in the lipoprotein metabolism.

Gene	Human			Mouse		
	Chromosomal localisation[1]	Gene structure[2]	References	Chromosomal localisation[3]	Gene structure[2]	References
APOA2	1q21-q23	+	Tsao et al., 1985	1	+	Higuchi et al., 1991
APOB	2p24-p23	+	Blackhart et al., 1986	12		
APOA1	11q23-q24	+	Karathanasi et al., 1985	9	+	Januzzi et al., 1992
APOC3	11q23-qter	+	Protter et al., 1984	9	+	Januzzi et al., 1992
APOA4	11q23-qter	+	Karathanasi et al., 1986	9	+	Williams et al., 1986
APOE	19q13.2	+	Das et al., 1985	7	+	Horiuchi et al., 1989
APOC1	19q13.2	+	Lauer et al., 1988	7	+*	Hoffer et al., 1993c
APOC2	19q13.2	+	Wei et al., 1985; Das et al., 1987	7	+*	Hoffer et al., 1993b
LDLR	19p13.3	+	Südhof et al., 1985	9	+*	Hoffer et al., 1993a
LRP	12q13-q14					
LPL	8p22	+	Deep et al., 1989; Kirschener et al., 1989	8	+	Zechner et al., 1991
LCAT	16q22.1	+	McLean et al., 1986			
CETP	16q13	+	Agellon et al., 1990			
HTGL	15q21-q23	+	Cai et al., 1989	9		

[1]According to Human Gene Mapping 11 (1991).
[2]Isolated genes are indicated by +.
[3]According to Hillyard et al., 1993.
*Described in this thesis.

also be conserved in humans (see below).

Linkage analysis can be used to identify genes for genetic disorders in which no biochemical defect or gene can be identified. In human, genetic markers are tested for cosegregation with the disease phenotype (monogenetic) in families. The development of highly polymorphic simple sequence repeats has been successful for the identification of loci causing human disease (Wijmenga et al., 1990). However, this approach is difficult to use in human studies involving multifactorial disorders such as atherosclerosis. The multiple segregating loci contributing to the disease are difficult to identify. Therefore, animal models are used for the identification of complex genetic traits.

One approach of identifying these genetic factors is called the quantitative trait locus (QTL) mapping. QTL mapping (for a review, see Warden et al., 1992) requires inbred strains and a highly characterized map of the genome. The development of the highly polymorphic simple sequence repeats made it feasible to use QTL mapping for the identification of multiple loci. For QTL mapping two different inbred strains are crossed to produce F_2 backcross progeny. These progeny are individually phenotyped for the trait of interest and genotyped for markers in the genome. Subsequently, statistical analysis are used to test for associations. QTL mapping is a quick and cheap alternative but only dominant traits can be mapped and there is no permanent material. QTL mapping resulted in the mapping of loci contributing to diabetes in the mouse (Todd et al., 1991).

Another approach is to use recombinant inbred (RI) strains (Justice et al., 1992) instead of F_2 backcross progeny. RI strains make it possible to analyze both dominant and recessive disease loci. Two inbred strains which differ for the trait of interest, are crossed and a series of new homozygous recombinant inbred strains is created from their progeny by 20 generations of inbreeding. A disadvantage of this approach is that the production of RI strains is time consuming. Each new recombinant inbred strain consists of a unique mixture of genes in a homozygous state derived from the two parental strains. The set of recombinant inbred strains permits rapid linkage analysis because alleles of linked genes tend to be linked in the same combination in the RI strains, whereas unlinked genes are randomized. Another advantage of the use of RI strains is that they provide an unlimited source of material for further studies.

QTL mappping depends on the variability of lipoproteins between inbred strains and therefore characterization of the lipoprotein metabolism is required. Among various strains of inbred mice, there is heterogeneity in the sizes, plasma concentrations and

lipid compositions of the lipoproteins (Jiao et al., 1990a, 1990b). It was also shown that there were strain-related differences with respect to mRNA quantities of genes involved in the lipoprotein metabolism (Lusis et al., 1987; Srivastava et al., 1991). These differences in quantitative traits between different inbred strains are used to study the multifactorial genetic influences on the lipoprotein metabolism.

The first approach in identifying genes involved in the variability of lipoproteins in the mouse is called the "candidate gene" approach. For this, polymorphisms in genes, involved in the human lipoprotein metabolism, are tested for cosegregation with specific quantitative traits. This approach has been used to identify genes controlling the structure of mouse lipoproteins. For instance, comparison of HDL sizes and HDL-cholesterol concentrations in plasma of several inbred strains of mice showed dichotomous distributions due to different isoforms of apoA1 and apoA2 (Ben-Zeev et al., 1983; Lusis et al., 1983, 1987). A problem of the "candidate gene" approach is that the "candidate gene" must be identified before one can test its involvement in the variability.

Therefore QTL mapping has been used as an approach to determine genetic factors influencing the variability of lipoproteins. Recombinant inbred strains were used to map the *Ath1* locus which is correlated with susceptibility to atherosclerosis (Paigen et al., 1985). Studying polymorphisms affecting HDL resulted in the identification of new additional loci involved in the heterogeneity of these lipoproteins in the mouse. When mice are fed an atherogenic diet, HDL-cholesterol decreased by 50% in susceptible strains whereas in the resistant strains the HDL-cholesterol levels remained unchanged. The frequency of lesion formation differs considerably between different strains fed the same atherogenic diet and is correlated with HDL-cholesterol levels (Paigen et al., 1985; Ishida et al., 1991). Variations in HDL-cholesterol levels and susceptibility to diet-induced atherosclerosis in mice appear to be determined by the same gene, named *Ath1* (Paigen et al., 1987; LeBoeuf et al., 1990). This new gene, maps to a region near the structural gene for apoA2 on mouse chromosome 1. This gene is distinct from previously known genes in this region. In addition, two other new loci, *Ath2* and *Ath3*, were also found to be linked to atherosclerosis and HDL phenotypes (Paigen et al., 1989; Stewart-Phillips et al., 1989). In the above mentioned studies, phenotypes are used to detect the underlying genes which are involved in the variability of lipoproteins.

4.3 Transgenic mouse models

Biochemical studies with mice have shown that there are differences in the relative amounts of the density classes of plasma lipoproteins between mice and humans. In mice, HDL is the predominant lipoprotein in the plasma compared to LDL in man. Despite this, when mice are fed a diet comparable with the human diet (atherogenic diet), their lipoprotein profile approaches that of humans. Atherogenic diet consumption results in an increase in total plasma cholesterol which could be attributed to an increase in VLDL/LDL-cholesterol (Paigen et al., 1985). An more elegant approach to influence the lipoprotein metabolism of the mouse is by the use of transgenesis and gene targeting. Transgenesis is a recently developed technology which makes it possible to introduce foreign DNA randomly into the mouse genome (Palmiter and Brinster, 1986; Jaenisch, 1988). These mouse models can be used to study the gene expression and the effects of overexpression of a transgene in the mouse. These animal models cannot be used to study the consequence of a decreased amount or total absence of the expression of a gene. Other technologies, like antisense expression vectors, can only be used to study the absence of expression of a gene (Weintraub, 1990). A more reliable technology is targeted mutagenesis (gene targeting), allowing the introduction of defined changes into any endogenous mouse gene in embryonic stem cells (Capecchi, 1989; Pascoe et al., 1992; Smithies, 1993). Gene targeting can be used to produce animal models of any characterized monogenic human disease or to study the function of new genes by the creation of null alleles (knock-out) for the appropriate gene. The effects of defined mutations can be studied against a homogeneous genetic background and under controlled environmental conditions. Both the genetic background as well as the environmental conditions can be adjusted as much as the researchers wishes. These technologies together, transgenesis and gene targeting, make it possible to study several aspects of gene expression in the mouse. Table 3 shows the transgenic mouse models for the lipoprotein metabolism and atherosclerosis available at present.

Transgenic mice can be used to determine the effects of overexpression of a particular gene on the lipoprotein metabolism. The overexpression of genes can result in a reduction of plasma lipoproteins and protection against diet-induced atherosclerosis. This effect has been found for human APOE, LDLR and APOA1 transgenic mouse models (Yokode et al., 1990; Rubin et al., 1991; Shimano et al., 1992). The expression of the LDL receptor in the liver of transgenic mice was used to study whether overexpression would increase the clearance of LDL from the

Table 3. Transgenic mouse models for studying the lipoprotein metabolism and atherosclerosis.

Gene	Phenotype[1]		References
	Transgenic	Knock-out	
APOA1	HDL↑		Walsh et al., 1989; Chajek-Shaul et al., 1991; Rubin et al., 1991; Golder-Novoselsky et al., 1992; Walsh et al., 1993; Hayek et al., 1993
APOA1/CETP	HDL-C↓		Hayek et al., 1992
APOA1/APOA2	HDL-C ↑		Schultz et al., 1992
APOA2	small HDL		Schultz et al., 1992; Warden et al., 1993
APOB		TG↓, CHOL↓	Homanics et al., 1993
APOC1	TG↑		Simonet et al., 1991
APOC3	TG↑↑		Ito et al., 1990; Aalto-Setälä et al., 1992
APOE	VLDL↓, LDL↓	CHOL↑↑, TG↑	Simonet et al., 1990, 1993; Smith et al., 1990; Piedrahita et al., 1992; Plump et al., 1992; Zhang et al., 1992; Popko et al., 1993
APOE3-Leiden	TG↑, CHOL↑		van den Maagdenberg et al., 1993
APOE/APOC1	Normal		Simonet et al., 1991
LDLR	LDL↓	IDL↑, LDL↑	Hofmann et al., 1988, 1990; Pathak et al., 1990; Yokode et al., 1990, 1992; Ishibashi et al., 1993
LRP		-*	Herz et al., 1992
CETP	HDL↓		Agellon et al., 1991; Jiang et al. 1992

[1] TG: triglycerides; CHOL/C: cholesterol.
* No homozygotes.

plasma. This was the first study with transgenic mice which showed that human and mouse lipoprotein metabolism are compatible, i.e. the human LDL receptor binds and internalizes mouse lipoproteins containing apoB and apoE. The overexpression of human LDL receptors resulted in a rapid clearance of LDL (Hofmann et al., 1988, 1990). In addition, the overexpression of the LDLR gene made it also possible to study the tissue-specific sorting of the LDL receptor. This has not been possible before since the number of receptors was below the detection limit in normal mice (Pathak et al., 1990).

In man, the study of mutant genes is often difficult due to additional factors influencing the expression of atherosclerosis. The use of a transgenic animal model offers the possibility to study the mutation in a homogenous environmental and genetic background. The overexpression of a mutant gene, apoE3-Leiden, in a transgenic mouse model resulted in an elevated lipoprotein plasma levels of VLDL and LDL (van den Maagdenberg et al., 1993). This transgenic mouse can serve as a model to elucidate additional factors influencing the plasma lipoproteins and for the etiology of familial dysbetalipoproteinemia.

Studies with transgenic animals can also be informative on tissue-specific regulation of the human gene involved. Transgenic mice can show different levels of expression of the construct involved which depends on the number of constructs which have been integrated and on the position in the genome. The expression of genes is highly variable and unpredictable unless locus control regions (LCR) are included in the constructs used. Studies showed that constructs containing the LCRs had all the regulatory elements involved in the expression of the gene involved (Constantoulakis et al., 1991). An example of an incomplete construct is the study with a transgenic mouse model for apoE containing only the 5' regulatory elements. This construct did not show expression in the liver which is the major site of apoE synthesis. Other studies using different constructs with the human APOE gene showed that the potential liver-specific enhancer element is located between the APOC1 and pseudo APOC1 genes. These constructs have been very useful to study the tissue-specific regulatory elements involved in the APOE gene expression (Smith et al., 1990; Simonet et al., 1990, 1991, 1993).

Gene targeting can be used to study the effect of a deficiency for the gene of interest. Inactivation of a gene (gene knock-out) is the most straigth forward approach to make a defect although the introduction of more subtle mutations are also possible. At present, only *Apoe*, *Apob*, *Ldlr*, and a *Lrp* knockout mice have been made

(Piedrahita et al., 1992; Zhang et al., 1992; Plump et al., 1992; Homanics et al., 1993; Ishibashi et al., 1993; Herz et al., 1992). *Apoe*-deficient mice showed a phenotype with spontaneous high cholesterol levels and atherosclerotic lesions which give them a unique value. These mice are a reproducible source of atherosclerotic animals which can be used for studies on the progression of the disease and influences of environmental factors without interference with an atherogenic diet.

For studying the lipoprotein metabolism and atherosclerosis, transgenic or gene targeted mice are preferable to mice with diet-induced atherosclerosis. These technologies have been used to produce lines of transgenic mice which carry human genes involved in the lipoprotein metabolism or knock-out mice for specific genes (Table 3). The expression of human genes of interest and the consequence of overexpression of these genes on the lipoprotein metabolism have been studied in these mice. The use of transgenic animals will be very useful in studying multifactorial diseases in future. The lines of transgenic and knock-out mice make it possible to create crosses between different lines. An example is the APOA1/CETP transgenic mouse which has been created from an APOA1 and a CETP transgenic mmouse by crossbreeding. Another possibility is to make crosses with other inbred strains. These new mouse models make it possible to study the same transgene with another genetic background. These studies will be very useful for unravelling the genetic factors which are involved in the variability of lipoproteins.

5. OUTLINE OF THE THESIS

The insight in the chylomicron, VLDL, and LDL metabolism, and the relationship with atherosclerosis has increased largely due to human studies. But it also clearly revealed the shortcomings, mainly related to the complex genetic and/or environmental interactions. Natural animal models have proven to be of great value to complement the human studies. Recent developments in the technology of producing specific transgenic mouse models by conventional transgenesis and targeted mutagenesis, will give unforseen new possibilities. However, this also further emphasizes the necessity of detailed insight in the structure and function of the corresponding mouse genes involved.

Accordingly, this thesis has been focused to the study of the *Apoe-c1-c2* gene cluster (Chapter 2-5) and the *Ldlr* gene (Chapter 6). These genes have been chosen because

they play a central role in the metabolism of atherogenic lipoproteins. Of the four mentioned genes, only the mouse *Apoe* gene has already been isolated and characterized (Table 2).

We have isolated the mouse *Apoe-c1-c2* gene cluster and the exact location of the *Apoc1* and *Apoc2* genes was determined relatively to the *Apoe* gene. The genetic variability in this gene cluster between inbred strains of mice has been determined (Chapter 2). The *Apoc1* gene (Chapter 3), the *Apoc2* gene (Chapter 4) have been isolated and characterized in more detail on cDNA and genomic level. The expression of these genes in mice was also investigated. The second part of chapter 4 reports the detection of a larger transcript which has not been described sofar and showed evolutionary conservation. Further study on this transcript resulted in the isolation of cDNA clones encoded by a novel gene within the *Apoe-c1-c2* gene cluster. This new gene, the apolipoprotein C2 linked *(Acl)* gene, has been further characterized (Chapter 5). Finally, the characterization of the *Ldlr* gene on cDNA and genomic level is described in Chapter 6.

REFERENCES

Aalto-Setälä K, Fisher EA, Chen X, Chajek-Shaul T, Hayek T, Zechner R, Walsh A, Ramakrishnan R, Ginsberg HN, and Breslow JL (1992). Mechanism of hypertriglyceridemia in human apolipoprotein (Apo) CIII transgenic mice. J Clin Invest 90: 1889-1900.

Agellon LB, Quinet EM, Gillette TG, Drayna DT, Brown ML, and Tall RA (1990). Organization of the human cholesteryl ester transfer protein gene. Biochem 29: 1372-1376.

Agellon JB, Walsh A, Hayek T, Moulin P, Jiang XC, Shelanski SA, Breslow JL, and Tall AR (1991).Reduced high density lipoprotein cholesterol in human cholesteryl ester transfer protein transgenic mice. J Biol Chem 266: 10796-10801.

Armstrong ML, and Heistad DD (1990). Animal models of atherosclerosis. Atherosclerosis 85: 15-23.

Auwerx JH, Chait A, Wolfbauer G, and Deeb SS (1989). Involvement of second messenger in regulation of the low density lipoprotein receptor gene. Mol Cell Biol 9: 2298-2302.

Ben-Zeev O, Lusis AJ, LeBoeuf RC, Nikazy J, and Schotz MC (1983). Evidence for independent genetic regulation of heart and adipose lipoprotein activity. J Biol Chem 258: 13632-13636.

Blackhart BD, Ludwig EM, Pierotti VR, Caiatai L, Onasch MA, Wallis SC, Powell L, Pease R, Knott TJ, Chu ML, Mahley RW, Scott J, McCarthy BJ, and Levy-Wilson B (1986). Structure of the human apolipoprotein B gene. J Biol Chem 261: 15364-15367.

Blue ML, Williams DL, Zucker S, Khan A, and Blum CB (1983). Apolipoprotein E synthesis in human kidney, adrenal gland, and liver. Proc Natl Acad Sci USA 80: 283-287.

Breckenridge WC, Little JA, Steiner G, Chow A, and Poapst M (1978). Hypertriglyceridemia associated with deficiency of apolipoprotein C-II. N Engl J Med 298: 1265-1273.

Breslow JL (1988). Apolipoprotein genetic variation and human disease. Physiol Rev 68: 85 132.

Brook JD, Skinner M, Roberts SH, Rettig WJ, Almond JW, and Shaw DJ (1987). Further mapping of markers around the centromere of human chromosome 19. Genomics 1: 320-328.

Brown MS, and Goldstein JL (1986). A receptor-mediated pathway for cholesterol homeostasis. Science 232: 34-47.

Cai S-J, Wong DM, Chen S-H, and Chan L (1989). Structure of the human hepatic triglyceride lipase gene. Biochem 28: 8966-8971.

Camus M-C, Chapman MJ, Forgez P, and Laplaud PM (1983). Distribution and characterization of the serum lipoproteins and apoproteins in the mouse, Mus musculus. J Lipid Res 24: 1210-1228.

Capecchi M (1989). The new mouse genetics altering the genome by gene targeting. Trends Genet 5: 70-76.

Catapano AL, Kinnunen PKJ, Breckenridge WC, Gotto AM Jr, Jackson RL, Little JA, Smith LC, and Sparrow JT (1979). Lipolysis of apoC-II deficient very low density lipoproteins: enhancement of lipoprotein lipase action by synthenic fragments of apoC-II. Biochem Biophys Res Commun 89: 951-957.

Chajek-Shaul T, Hayer T, Walsh A, and Breslow JL (1991). Expression of the human apolipoprotein A-I gene in transgenic mice alters high density lipoprotein (HDL) particle size distribution and diminishes selective uptake of HDL cholesteryl esters. Proc Natl Acad Sci USA 88: 6731-6735.

Chen W-J, Goldstein JL, and Brown MS (1990). NPXY, a sequence often found in cytoplasmatic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. J Biol Chem 265: 3116-3123.

Cheng Q, Blackett P, Jackson KW, McConathy WJ, and Wang C-S (1990). C-terminal domain of apolipoprotein CII as both activator and competitive inhibitor of lipoprotein lipase. Biochem J 269: 403-407.

Constantoulakis P, Josephson B, Mangahas L, Papayannopoulou T, Enver T, Constantini F, and Stamatoyannopoulos (1991). Locus control region-Ax transgenic mice: a new model for studying the induction of fetal hemoglobin in the adult. Blood 77: 1326-1333.

Cornall RJ, Aitman TJ, Hearne CM, and Todd JA (1991). The generation of a library of PCR analyzed microsatellite variants for genetic mapping of the mouse genome. Genomics 10: 874-881.

Das HK, McPherson J, Bruns AP, Karathanasis SK, and Breslow JL (1985). Isolation, characterization and mapping to chromosome 19 of the human apolipoprotein E gene. J Biol Chem 260: 6240-6247.

Das HK, Jackson CL, Miller DA, Leff T, and Breslow JL (1987). The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron. J Biol Chem 262: 4787-4793.

Davis CG, Elhammer A, Russell DW, Schneider WJ, Kornfeld S, Brown MS, and Goldstein JL (1986). Deletion of clustered O-linked carbohydrates does not impair function of low density lipoprotein receptor in transfected fibroblasts. J Biol Chem 261: 2828-2838.

Davis CG, Goldstein JL, Südhof TC, Anderson RGW, Russell DW, and Brown MS (1987). Acid-dependent ligand dissociation and recycling of LDL receptor mediated by growth factor homology region. Nature 326:760-765.

Davison PJ, Norton P, Wallis SC, Gill L, Cook M, Williamson R, and Humphries SE (1986). There are two gene sequences for human apolipoprotein CI (APOCI) on chromosome 19, one of which is 4 kb from the gene for APOE. Biochem Biophys Res Commun 136: 876-884.

Dawson PA, Hofmann SL, Van der Westhuyzen DR, Südhof TC, Brown MS, and Goldstein JL (1988). Sterol-dependent repression of low density lipoprotein receptor promoter mediated by 16-base pair sequence adjacent to binding site for transcription factor Sp1. J Biol Chem 263: 3372-3379.

Deep SS, and Peng R (1989). Structure of the human lipoprotein lipase gene. Biochem 28: 4131-4135.

Eisenberg S (1990). Metabolism of apolipoproteins and lipoproteins. Curr Opin Lipidol 1: 205 215.

Esser V, Limbird LE, Brown MS, Goldstein JL, and Russell DW (1988). Mutational analysis of the ligand binding domain of the low density lipoprotein receptor. J Biol Chem 263: 13282-13290.

Fojo SS, Law SW, and Brewer HB Jr (1984). Human apolipoprotein C-II: complete nucleic acid sequence of preapolipoprotein C-II. Proc Natl Acad Sci USA 81: 6354-6357.

Golder-Novoselsky E, Forte TM, Nichols AV, and Rubin EM (1992). Apolipoprotein AI expression and

high density lipoprotein distribution in transgenic mice during development. J Biol Chem 267: 20787-20790.

Goldstein JL, and Brown MS (1989). Familial hypercholesterolemia. In: The metabolic basis of inherited disease (Scriver CR, Beaudet AL, Sly WS, and Valle D, Eds.). 6th ed McGraw-Hill, New York, pp. 1215-1250.

Gotto AM, Pownall HJ, and Havel JR (1986). Introduction to the plasma lipoproteins. Methods Enzymol 128: 3-40.

Havel RJ, and Kane JP (1989). Introduction: Structure and metabolism of plasma lipoproteins. In: The metabolic basis of inherited disease (Scriver CR, Beaudet AL, Sly WS, and Valle D, Eds.). 6th ed McGraw-Hill, New York, pp. 1129-1138.

Hayek T, Chajek-Shaul T, Walsh A, Agellon LB, Moulin P, Tall AR, and Breslow JL (1992). An interaction between the human cholesteryl ester transfer protein (CETP) and apolipoprotein A-I genes in transgenic mice results in a profound CETP-mediated depression of high density lipoprotein cholesterol levels. J Clin Invest 90: 505-510.

Herz J, Clouthier DE, and Hammer RE (1992). LDL receptor-related protein internalizes and degrades uPA-PAI-1 complexes and is essential for embryonic implantation. Cell 71: 411-421.

Higuchi K, Kitagawa K, Naiki H, Hanada K, Hosokawa M, and Takeda T (1991). Polymorphism of apolipoprotein A-II (apoA-II) among inbred strains of mice. Biochem J 279: 427-433.

Hillyard AL, Doolittle DP, Davisson MT, and Roderick TH (1993). Locus map of the mouse withcomparative map points of human on mouse. Mouse Genome 91: 15-39.

Hobbs HH, Russell DW, Brown MS, and Goldstein JL (1990). The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. Annu Rev Genet 24: 133-170.

Hobbs HH, Brown MS, and Goldstein JL (1992). Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Hum Mutation 1: 445-466.

Hoffer MJV, van Eck MM, Petrij F, van der Zee A, de Wit E, Meijer D, Grosveld G, Havekes LM, Hofker MH, and Frants RR (1993a). The mouse low density lipoprotein receptor gene: cDNA sequence and exon-intron structure. Biochem Biophys Res Commun 191:880-886.

Hoffer MJV, van Eck MM, Havekes LM, Hofker MH, and Frants RR (1993b). Structure and expression of the mouse apolipoprotein C2 gene. Genomics 17:45-51.

Hoffer MJV, van Eck MM, Havekes LM, Hofker MH, and Frants RR (1993c). The mouse apolipoprotein C1 gene: structure and expression. Genomics: in press.

Hofmann SL, Russell DW, Brown MS, Goldstein JL, and Hammer RE (1988). Overexpression of low density lipoprotein (LDL) receptor eliminates LDL from plasma in transgenic mice. Science 239: 1277-1281.

Hofmann SL, Eaton DL, Brown MS, McConathy WJ, Goldstein JL, and Hammer RE (1990). Overexpression of human low density lipoprotein receptors leads to accelerated catabolism of Lp(a) lipoprotein in transgenic mice. J Clin Invest 85: 1542-1547.

Homanics GE, Smith TJ, Zhang SH, Lee D, Young SG, and Maeda N (1993). Targeted modification of the apolipoprotein B gene results in hypobetalipoproteinemiaand developmental abnormalities in mice. Proc Natl Acad Sci USA 90: 2389-2393.

Horiuchi K, Tajima S, Menju M, and Yamamoto A (1989). Structure and expression of mouse apolipoprotein E gene. J Biochem 106: 98-103.

Hospattankar AV, Fairwell T, Ronan R, and Brewer HB Jr (1984). Amino acid sequence of human plasma apolipoprotein C-II from normal and hyperlipoproteinemic subjects. J Biol Chem 259: 318-322.

Innerarity TL, Weisgraber KH, Arnold KS, Rall SC Jr, and Mahley RW (1984). Normalization of receptor binding of apolipoprotein E2: Evidence for modulation of the binding site conformation. J Biol Chem 259: 7261-7267.

Ishida BY, Blanche PJ, Nichols AV, Yashar Y, and Paigen B (1991). Effects of atherogenic diet consumption on lipoproteins in mouse strains C57BL/6 and C3H. J Lipid Res 32: 559-568.

Ishibashi S, Brown MS, Goldstein JL, Gerard RD, Hammer RE, and Herz J (1993). Hypercholesterolemia

in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery. J Clin Invest 92: 883-893.

Ito Y, Azrolan N, O'Connell A, Walsh A, and Breslow JL (1990). Hypertriglyceridemia as a result of human apo CIII gene expression in transgenic mice. Science 249: 790-793.

Jackson CL, Bruns GAP, and Breslow JL (1984). Isolation and sequence of a human apolipoprotein CII cDNA clone and its use to isolate and map to human chromosome 19 the gene for apolipoprotein CII. Proc Natl Acad Sci USA 81: 2945-2949.

Jackson RL, Sparrow JT, Baker HN, Morrisett J, Taunton OD, and Gotto AM Jr (1974). The primary structure of apolipoprotein-serine. J Biol Chem 249: 5308-5313.

Jackson RL, Baker HN, Gilliam EB, and Gotto AM Jr (1977). Primary structure of very low density apolipoprotein C-II of human plasma. Proc Natl Acad Sci USA 74: 1942-1945.

Jaenisch R (1988). Transgenic animals. Science 240: 1468-74.

Januzzi JL, Azrolan N, O'Connell A, Aalto-Setälä K, and Breslow JL (1992). Characterization of the mouse apolipoprotein *Apoa-1/Apoc-3* gene locus: genomic, mRNA, and protein sequences with comparisons to other species. Genomics 14: 1081-1088.

Jiang XC, Agellon LB, Walsh A, Breslow JL, and Tall A (1992). Dietary cholesterol increases transcription of the human cholesteryl ester transfer protein gene in transgenic mice. J Clin Invest 90: 1290-1295.

Jiao S, Cole TG, Kitchens RT, Pfleger B, and Schonfeld G (1990a). Genetic heterogeneity of lipoproteins in inbred strains of mice: analysis by gel-permeation chromatography. Metabolism 39: 155-160.

Jiao S, Cole TG, Kitchens T, Pfleger B, and Schonfeld G (1990b). Genetic heterogeneity of plasma lipoproteins in the mouse: control of low density lipoprotein particle sizes by genetic factors. J Lipid Res 31: 467-477.

Justice MJ, Jenkins NA, and Copeland NG (1992). Recombinant inbred mouse strains: models for disease study. Trends Biotech 10: 120-126.

Karathanasis SK, Zannis VI, and Breslow JL (1985). Isolation and characterization of the human apolipoprotein A-I gene. Proc Natl Acad Sci USA 80: 6147-6151.

Karathanasis SK, Oettgen P, Haddad IA, and Antonarakis SE (1986). Structure, evolution, and polymorphisms of the human apolipoprotein A4 gene (APOA4). Proc Natl Acad Sci USA 83: 8457-8461.

Kinnunen PKJ, Jackson RL, Smith LC, Gotto AM Jr, and Sparrow JT (1977). Activation of lipoprotein lipase by native and synthetic fragments of human plasma apolipoprotein C-II. Proc Natl Acad Sci USA 74: 4848-4851.

Kirchgessner TG, Chaut JC, Heinzmann C, Eitienne J, Guflhot S, Svenson K, Ameis D, Pilon C, D'Auriol L, Andalibi A, Schotz MC, Galifert F, and Lusis AJ (1989). Organization of the human lipoprotein lipase gene and evolution of the lipase family. Proc Natl acad Sci USA 86: 9647-9651.

Knott TJ, Robertson ME, Priestley LM, Wallis S, and Scott J (1984). Characterization of mRNAs encoding the precursor for human apolipoprotein CI. Nucl Acids Res 12: 3909-3915.

Kowal RC, Herz J, Weisgraber KH, Mahley RW, Brown MS, and Goldstein JL (1990). Opposing effects of apolipoproteins E and C on lipoprotein binding to low density lipoprotein receptor-related protein. J Biol Chem 265: 10771-10779.

LaRosa JC, Levy RI, Herbert P, Lux SE, and Frederickson DS (1970). A specific apoprotein activator for lipoprotein lipase. Biochem Biophys Res Commun 41: 57-62.

Lauer SJ, Walker D, Elshourbagy NA, Reardon CA, Levy-Wilson B, and Taylor JM (1988). Two copies of the human apolipoprotein C-I gene are linked closely to the apolipoprotein E gene. J Biol Chem 263: 7277-7286.

LeBoeuf RC, Puppion DL, Schumaker VN, and Lusis AJ (1983). Genetic control of lipid transport in mice. I. Structural properties and polymorphisms of plasma lipoproteins. J Biol Chem 258: 5063-5070.

LeBoeuf RC, Doolittle MH, Montcalm A, Martin DC, Reue K, and Lusis AJ (1990). Phenotypic characterization of the Ath-1 gene controlling high density lipoprotein levels and susceptibility to

atherosclerosis. J Lipid Res 31: 91-101.

Li W-H, Tanimura M, Luo C-C, Datta S, and Chan L (1988). The apolipoprotein multigene family: biosynthesis, structure, structure-evolution relationships, and evolution. J Lipid Res 29: 245-271.

Lindgren V, Luskey KL, Russell DW, and Francke U (1985). Human genes involved in cholesterol metabolism: Chromosomal mapping of the loci for the low density lipoprotein receptor and 3-hydroxy-3-methylglutaryl-coenzyme A reductase with cDNA probes. Proc Natl Acad Sci USA 82: 8567-8571.

Lusis AJ, Taylor BA, Wangenstein RW, and LeBoeuf RC (1983). Genetic control of lipid transport in mice. II. Genes controlling structure of high density lipoproteins. J Biol Chem 258: 5071-5078.

Lusis AJ, and LeBoeuf RC (1986). Genetic control of plasma lipid transport: mouse model. Methods in Enzymol 128: 877-894.

Lusis AJ, Taylor BA, Quon D, Zollman S, and LeBoeuf RC (1987). Genetic factors controlling structure and expression of apolipoproteins B and E in mice. J Biol Chem 262: 7594-7604.

Lusis, AJ (1988). Genetic factors affecting blood lipoproteins: the candidate gene approach. J Lipid Res 29: 397-429.

Lusis AJ, and Sparks RS (1989). Chromosomal organization of genes involved in plasma lipoprotein metabolism: human and mouse "fat maps". In: Genetic factors in atherosclerosis: approaches and model systems (Lusis AJ, and Sparkes SR, Eds.). Monogr Hum Genet, Basel Karger. vol 12, pp. 79-94.

Mahley RW, Innerarity TL, Rall SC Jr, and Weisgraber KH (1984). Plasma lipoproteins: apolipoprotein structure and function. J Lipid Res 25: 1277-1294.

Mahley RW, and Rall SC Jr (1989). Type III hyperlipoproteinemia (dysbetalipoproteinemia): The role of apolipoprotein E in normal and abnormal lipoprotein metabolism. In: The metabolic basis of inherited disease (Scriver CR, Beaudet AL, Sly WS, and Valle D, Eds). 6th ed McGraw-Hill, New York, pp. 1195-1213.

Marotti KR, Castle CK, Boyle TP, Lin AH, Murray RW, and Melchior GW (1993). Severe atherosclerosis in transgenic mice expressing simian cholesteryl ester transfer protein. Nature 364: 73-75.

Mazzone T, Basheeruddin K, Ping L, and Schick C (1990). Relation of growth- and sterol related regulatory pathways for low density lipoprotein receptor gene expression. J Biol Chem 265: 5145-5149.

McLean JW, Elshourbagy NA, Chang DJ, Mahley RW, and Taylor JM (1984). Human apolipoprotein E mRNA. cDNA cloning and nucleotide sequencing of a new variant. J Biol Chem 259: 6498-6504.

McLean J, Wion K, Drayna D, Fielding C, and Lawn R (1986). Human lecithin-cholestrol acyltransferase gene: complete gene sequence and sites of expression. Nucl Acid Res 14: 9397-9406.

Morrisett JD, Kim H-S, Patsch JR, Datta SK, and Trentin JJ (1982). Genetic susceptibility and resistance to diet-indiced atherosclerosis and hyperlipoproteinemia. Ateriosclerosis 2: 312-324.

Myklebost O, Williamson B, Markham AF, Mycklebost SR, Rogers J, Woods DE, and Humphries SE (1984). The isolation and characterization of cDNA clones for human apolipoprotein CII. J Biol Chem 259: 4401-4404.

Myklebost O, and Rogne S (1986). The gene for human apolipoprotein CI is located 4.3 kb away from the apolipoprotein E gene on chromosome 19. Hum Genet 73: 286-289.

Myklebost O, and Rogne S (1988). A physical map of the apolipoprotein gene cluster on human chromosome 19. Hum Genet 78: 241-247.

Nadeau JH and Taylor BA (1984). Lengths of chromosomal segments conserved since divergence of man and mouse. Proc Natl Acad Sci USA 81: 814-818.

Nishina PM, Verstuyft J, and Paigen B (1990). Synthetic low and high fat diets for the study of atherosclerosis in the mouse. J Lipid Res 31: 859-869.

Overturf ML, and Loose-Mitchell DS (1992). In vivo model systems: the choise of the experimental animal models for analysis of lipoproteins and atherosclerosis. Curr Opin Lipidol 3: 179-185.

Paigen B, Morrow A, Brandon C, Mitchell D, and Holmes P (1985). Variation in susceptibility to atherosclerosis among inbred strains of mice. Atherosclerosis 57: 65-73.

Paigen B, Mitchell D, Reue K, Morrow A, Lusis AJ, and Leboeuf RC (1987). Ath-I, a gene determining atherosclerosis susceptibility and hig density lipoprotein levels in mice. Proc Natl Acad Sci USA 84:

3763-3767.

Paigen B, Nesbitt MN, Mitchell D, Albee D, and LeBoeuf RC (1989). Ath2, a second gene determining atherosclerosis susceptibility and high density lipoprotein levels in mice. Genetics 122: 163-168.

Paik YK, Chang DJ, Reardon CA, Davis GE, Mahley RW, and Taylor JM (1985). Nucleotide sequence and structure of the human apolipoprotein E gene. Proc Natl Acad Sci USA 82: 3445-3449.

Palmiter RD, and Brinster RL (1986). Germ-line transformation of mice. Ann Rev Genet 20: 465-499.

Pascoe WS, Kemler R, and Wood SA (1992). Genes and functions: trapping and targetting in embryonic stem cells. Biochim Biophys Acta 1114: 209-221.

Pathak RK, Yokode M, Hammer RE, Hofmann SL, Brown MS, Goldstein JL, and Anderson RGW (1990). Tissue-specific sorting of the human LDL receptor in polarized epithelia of transgenic mice. J Cell Biol 111: 347-359.

Piedrahita JA, Zhang SH, Hagaman JR, Oliver PM, and Maeda N (1992). Generation of mice carrying a mutant apolipoprotein E gene inactivated by gene targetting in embryonic stem cells. Proc Natl Acad Sci USA 89: 4471-4475.

Plump AS, Smith JD, Hayek T, Aalto-Selätä K, Walsh A, Verstuyft JG, Rubin EM, and Breslow JL (1992). Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. Cell 71: 343-353.

Poensgen J (1990). Apolipoprotein C-I inhibits the hydrolysis by phosphlipase A_2 of phospholipids in liposomes and cell membranes. Biochim Biophys Acta 1042: 188-192.

Popko B, Goodrum JF, Bouldin TW, Zhang SH, and Maeda N (1993). Nerve regeneration occurs in the absence of apolipoprotein E in mice. J Neurochem 60: 1155-1158.

Protter AA, Levy-Wilson B, Miller J, Bencen G, White T, and Seilhamer JJ (1984). Isolation and sequence analysis of the human apolipoprotein CIII gene and the intergenic region between the apo AI and apo CIII genes. DNA 3: 449-456.

Raisonnier A (1991). Duplication of the apolipoprotein C-I gene occurred about forty million years ago. J Mol Evol 32: 211-219.

Rall SC, Weisgraber KH, and Mahley RW (1982). Human apolipiprotein E. The complete amino acid sequence. J Biol Chem 257: 4171-4178.

Rapacz J, Hasler-Rapacz J, Taylor KM, Checovich WJ, and Attle AD (1986). Lipoprotein mutations in pigs are associated with elevated plasma cholesterol and atherosclerosis. Science 234: 1573-1577.

Reue KL, Quon DH, O'Donnell KA, Dizikes GJ, Fareed GC, and Lusis AJ (1984). Cloning and regulation of messenger RNA for mouse apolipoprotein E. J Biol Chem 259: 2100-2107.

Reue KL, Warden CH, and Lusis AJ (1990). Animal models for genetic lipoprotein disorders. Curr Opin Lipidol 1: 143-150.

Roberts A, and Thompson JS (1976). Inbred mice and their hybrids as an animal model for atherosclerosis research. Adv Exp Med Biol 67: 313-327.

Rubin EM, Krauss RM, Spangler EA, Verstuyft JG, and Clift SM (1991). Inhibition of early atherogenesis in transgenic mice by human apolipoprotein AI. Nature 353: 265-267.

Russell DW, Yamamoto T, Schneider WJ, Luskey KL, Brown MS, and Goldstein JL (1984). Domain map of the LDL receptor: sequence homology with the epidermal growth factor precursor. Cell 37: 577-585.

Russell DW, Brown MS, and Goldstein JL (1989). Different combinations of cysteine-rich repeats mediate binding of low density lipoprotein receptor to two different proteins. J Biol Chem 264: 21682-21688.

Saunders AM, and Seldin MF (1990). A molecular linkage map of mouse chromosome 7. Genomics 8: 525-535.

Schneider WJ, Beisiegel U, Goldstein JL, and Brown MS (1982). Purification of the low density lipoprotein receptor, an acidic glycoprotein of 164,00 molecular weigth. J Biol Chem 257: 2664-2673.

Schultz JR, Gong EL, McCall MR, Nichols AV, Clift SM, and Rubin EM (1992). Expression of humans apolipoprotein A-II and its effect on high density lipoproteins in transgenic mice. J Biol Chem 267: 21630-21636.

Sehayek E, and Eisenberg S (1991). Mechanisms of inhibition by apolipoprotein C of apolipoprotein E

dependent cellular metabolism of human triglyceride-rich lipoproteins through the low density lipoprotein receptor pathway. J Biol Chem 266: 18259-18267.

Sharpe CR, Sidoli A, Shelley CS, Lucero MA, Shoulders CC, and Baralle FE (1984). Human apolipoproteins AI, AII, CII, and CIII, cDNA sequences and mRNA abundance. Nucl Acids Res 12: 3917-3932.

Shimano H, Yamada N, Katsuki M, Shimada M, Gotoda T, Harade K, Murase T, Fukazawa C, Takaku F, and Yazki Y (1992). Overexpression of apolipoprotein E in transgenic mice: marked reduction in plasma lipoproteins except high density lipoprotein and resistance against diet-induced hypercholesterolemia. Proc Natl Acad Sci USA 89: 1750-1754.

Shulman RS, Herbert PN, Wehrly K, and Frederickson DS (1975). The complete amino acid sequence of C-I (apoLP-Ser), an apolipoprotein from human very low density lipoproteins. J Biol Chem 250: 182-190.

Simonet WS, Bucay N, Lauer SJ, Wirak DO, Stevens ME, Weisgraber KH, Pitas RE, and Taylor JM (1990). In the absence of a downstream element, the apolipoprotein E gene is expressed at high levels in kidney of transgenic mice. J Biol Chem 265: 10809-10812.

Simonet WS, Bucay N, Pitas RE, Lauer SJ, and Taylor JM (1991). Multiple tissue-specific elements control the apolipoprotein E/C-I gene locus in transgenic mice. J Biol Chem 265: 8651-8654.

Simonet WS, Bucay N, Lauer SJ, and Taylor JM (1993). A Far-downstream hepatocyte-specific control region direct expression of the linked human apolipoprotein E and CI genes in transgenic mice. J Biol Chem 268: 8221-8229.

Smit M, van der Kooij-Meijs E, Frants RR, Havekes L, Klasen EC (1988a). Apolipoprotein gene cluster on chromosome 19. Hum Genet 78: 90-93.

Smit M, de Knijff P, Rosseneu M, Bury J, Klasen E, Frants R, and Havekes LM (1988b). Apolipoprotein E polymorphism in the Netherlands and its effect on plasma lipid and apolipoprotein levels. Hum Genet 80: 287-292.

Smith JD, Plump AS, Hayek T, Walsh A, and Breslow JL (1990a). Accumulation of human apolipoprotein E in the plasma of transgenic mice. J Biol Chem 265: 14709-14712.

Smith JR, Osborne TF, Goldstein JL, and Brown MS (1990b). Identification of nucleotides responsible for enhancer activity of sterol regulatory elements in low density lipoprotein receptor gene. J Biol Chem 265: 2306-2310.

Smithies O (1993). Animal models of human genetic diseases. Trens Genet 9: 112-116.

Soutar AK, Sigler GF, Smith LC, Gotto AM, and Sparrow JT (1978). Lecithin:cholesterol acyltransferase activation and lipid binding by synthetic fragments of apolipoprotein C-I. Scand J Clin Lab Invest 38: 53-58.

Soutar AK, Garner CW, Baker HN, Sparrow JT, Jackson RL, Gotto AM, and Smith LC (1975). Effect of the human plasma apolipoproteins and phosphatidylcholine acyl donor on the activity of lecithin: cholesterol acyltransferase. Biochem 14: 3057-3064.

Soutar AK (1992). Familial hypercholesterolemia and LDL receptor mutations. J Int Med 231: 633-641.

Srivastava RA, Jiao S, Tang J, Pfleger BA, Kitchens RT, and Schonfeld G (1991). In vivo regulation of low-density lipoprotein receptor and apolipoprotein B gene expressions by dietary fat and cholesterol in inbred strains of mice. Biochim Biophys Acta 1086: 29-43.

Stewart-Phillips JL, Lough J, and Skamene E (1989). ATH-3, a new gene for atherosclerosis in the mouse. Clin Invest Med 12: 121-126.

Stewart-Phillips JL, and Lough J (1991). Pathology of atherosclerosis in cholesterol-fed, susceptible mice. Atherosclerosos 90: 211-218.

Südhof TC, Goldstein JL, Brown MS, and Russell DW (1985). The LDL receptor gene: a mosaic of exons shared with different proteins. Science 228: 815-822.

Südhof TC, Van Der Westhuyzen DR, Goldstein JL, Brown MS, and Russell DW (1987a). Three indirect repeats and a TATA-like sequence are required for regulated expression of the human low density lipoprotein receptor gene. J Biol Chem 262: 10773-10779.

Südhof TC, Russell DW, Brown MS, and Goldstein JL (1987b). 42 bp element from LDL receptor gene confers end-product repression by sterols when inserted into viral TK promoter. Cell 48: 1061-1069.

Tall AR (1992). Metabolic and genetic control of HDL cholesterol levels. J Int Med 231: 661 668.

Taylor JM, Simonet WS, Bucay N, Lauer SJ, and de Silva HV (1991). Expression of the human apolipoprotein E/apolipoprotein C-I locus in transgenic mice. Curr Opin Lipidol 2: 73-80.

Todd JA, Aitman TJ, Cornall RJ, Ghosh S, Hall JRS, Hearne CM, Knight AM, Love JM, McAleer MA, Prins J-B, Rodrigues N, Lathrop M, Pressey A, Delarato NH, Peterson LB, and Wicker LS (1991). Genetic analysis of autoimmune type 1 diabetes mellitus in mice. Nature 351: 542-547.

Tolleshaug H, Goldstein JL, Schneider WJ, and Brown MS (1982). Posttranslational processing of the LDL receptor and its genetic disruption in familial hypercholesterolemia. Cell 30: 715-724.

Tsao Y-K, Wei C-F, Robberson DL, Gotto AM, and Chan L (1985). Isolation and characterization of the human apolipoprotein A-II gene. J Biol Chem 260: 15222-15231.

Utermann G, Steinmetz A, and Weber W (1982). Genetic control of human apolipoprotein E polymorphism: comparison of one- and two-dimensial techniques of isoprotein analysis. Hum Genet 60: 344-351.

Utermann, G (1985). Genetic polymorphism of apolipoprotein E - Impact on plasma lipoprotein metabolism. In: Diabetes, obesity and hyperlipidemias (Crepaldi G, Tiengo A, and Baggio G, Eds.). Amsterdam, NY: Elsevier. 3: 1-28.

van den Maagdenberg AMJM, Hofker MH, Krimpenford PJA, de Bruijn I, van Vlijmen B, van der Boom H, Havekes LM, and Frants RR (1993). Transgenic mice carrying the apolipoprotein E3-Leiden gene exhibit hyperlipoproteinemia. J Biol Chem 268: 10540-10545.

Wade DP, Knight BL, and Soutar AK (1989). Regulation of low-density-lipoprotein-receptor mRNA by insulin in human hepatoma Hep G2 cells. Eur J Biochem 181: 727-731.

Wallis SC, Rogne S, Gill L, Markham A, Edge M, Woods D, Williamson R, and Humphries SE (1983). The isolation of cDNA clones for human apolipoprotein E and the detection of apo E RNA in hepatic and extra-hepatic tissues. EMBO J 2: 2369-2373.

Walsh A, Ito Y, and Breslow JL (1989). High levels of human apolipoprotein A-I in transgenic mice result in increased plasma levels of small high density lipoprotein (HDL) particles comparable to human HDL_3. J Biol Chem 264: 6488-6494.

Walsh A, Azrolan N, Wang K, Marcigliano A, O'Connell A, and Breslow JL (1993). Intestinal expression of the human apoA-I gene in transgenic mice is controlled by a DNA region 3' to the gene in the promoter of the adjacent convergently transcribed apoC-III gene. J Lipid Res 34: 617-623.

Warden CH, Daluiski A, and Lusis AJ (1992). Identification of new genes contributing to atherosclerosis: the mapping of genes contributing to complex disorders in animal models. In: Molecular genetics of coronary artery disease. Candidate genes and processes in atherosclerosis. (Eds. Lusis AJ, Rotter JI, and Sparkes RS) Monogr Hum Genet, Basel Karger, vol 14, pp 419-441.

Warden CH, Hedrick CC, Qiao J-H, Castellani LW, and Lusis AJ (1993). Atherosclerosis in transgenic mice overexpressing apolipoprotein A-II. Science 261: 469-472.

Wei C-F, Tsao Y-K, Robberson DL, Gotto AM Jr, Brown K, and Chan L (1985). The structure of the human apolipoprotein C-II gene. J Biol Chem 260: 15211-15221.

Weintraub HM (1990). Antisense RNA and DNA. Scientific American 264: 34-40.

Weisgraber KH, Rall SC, Mahley RW, Milne RW, Marcel YL, and Sparrow JT (1986). Human apolipoprotein E. Determination of the heparin binding sites of apolipoprotein E3. J Biol Chem 261: 2068-2076.

Weisgraber KH, Mahley RW, Kowal RC, Herz J, Goldstein JL, and Brown MS (1990). Apolipoprotein C-I modulates the interaction of apolipoprotein E with β-migrating very low density lipoproteins (β-VLDL) and inhibits binding of β-VLDL to low density lipoprotein receptor-related protein. J Biol Chem 265: 22453-22459.

Wijmenga C, Frants RR, Brouwer OF, Moerer P, Weber JL, and Padberg GW (1990). Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. Lancet 336: 651-653.

Williams SC, Bruckheimer SM, Lusis AJ, LeBoeuf RC, and Kinniburgh AJ (1986). Mouse apolipoprotein A-IV gene: nucleotide sequence and induction by a high-lipid diet. Mol Cell Biol 6: 3807-3814.

Wilson C, Wardell MR, Weisgraber KH, Mahley RW, and Agard DA (1991). Three dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. Science 252: 8447-8450.

Windler E, and Havel RJ (1985). Inhibitory effect of C apolipoprotein from rats and humans on the uptake of triglyceride-rich lipoproteins and their remnants by the perfused rat liver. J Lipid Res 26: 556-565.

Yamamoto T, Davis CG, Brown MS, Schneider WJ, Casey ML, Goldstein JL, and Russell DW (1984). The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. Cell 39: 27-38.

Yamamoto T, Bischop RW, Brown MS, Goldstein JL, and Russell DW (1986). Deletion in cysteine-rich region of LDL receptor impedes transport to cell surface in WHHL rabbit. Science 232: 1230-1237.

Yokode M, Hammer SE, Ishibashi S, Brown MS, and Goldstein JL (1990). Diet-induced hypercholesterolemiain mice: prevention by overexpression of LDL receptors. Science 250: 1273-1275.

Yokodo M, Pathak RK, Hammer RE, Brown MS, Goldstein JL, and Anderson RGW (1992). Cytoplasmic sequence required for basolateral targetting of LDL receptors in livers of transgenic mice. J Cell Biol 117: 39-46.

Zannis VI, Just PW, and Breslow JL (1981). Human apolipoprotein E isoprotein subclasses are genetically determined. Am J Hum Genet 33: 11-24.

Zannis VI, Breslow JL, Utermann G, Mahley RW, Weisgraber KH, Havel RJ, Goldstein JL, Brown MS, Schonfeld G, Hazzard WR, and Blum C (1982). Proposed nomenclature of apoE isoproteins, apoE genotypes and phenotypes. J Lipid Res 23: 911-914.

Zannis VI, Kardassis D, Cardot P, Hadzopoulou-Cladaras M, Zanni EE, and Cladaras C (1992). Molecular biology of the human apolipoprotein genes: gene regulation and structure/function relationship. Curr Opin Lipidol 3: 96-113.

Zechner R, Newman TC, Steiner E, and Breslow JL (1991). The structure of the mouse lipoprotein lipase gene: a B1 repetitive element is inserted into the 3' untranslated region of the mRNA. Genomics 11: 62-76.

Zhang SH, Reddick RL, Piedrahita JA, Maeda N (1992). Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E. Science 258: 468-471.

CHAPTER 2

EVOLUTIONARY CONSERVATION OF THE MOUSE APOLIPOPROTEIN E-C1-C2 GENE CLUSTER: STRUCTURE AND GENETIC VARIABILITY IN INBRED MICE.

M.J.V. Hoffer, M.H. Hofker, M.M. van Eck, L.M. Havekes, and R.R. Frants.

Genomics 1993, 15: 62-67

Evolutionary Conservation of the Mouse Apolipoprotein *e–c1–c2* Gene Cluster: Structure and Genetic Variability in Inbred Mice

MARIËTTE J. V. HOFFER, MARTEN H. HOFKER,[1] MARGA M. VAN ECK, LOUIS M. HAVEKES,* AND RUNE R. FRANTS

*MGC-Department of Human Genetics, Leiden University, P.O. Box 9503, 2300 RA Leiden, The Netherlands; and *IVVO-TNO, Gaubius Laboratory, P.O. Box 430, 2300 AK Leiden, The Netherlands*

Received June 22, 1992, revised September 18, 1992

The human apolipoprotein E (APOE), APOC1, pseudo APOC1 (APOC1'), and APOC2 genes are clustered within 48 kb on the long arm of chromosome 19. A mouse *Apoe* cDNA probe was used to isolate overlapping cosmid clones from a cosmid library of the C57BL/Rij inbred mouse strain. These clones were investigated for the presence of the *Apoc1* and *Apoc2* genes by heterologous hybridization. Our results show that the *Apoe–c1–c2* gene cluster is conserved in the mouse. In line with evolutionary data, the mouse lacks the equivalent of APOC1'. These data were confirmed using a mouse *Apoc2* cDNA clone, and surprisingly the cDNA clone isolated here was 965 bp in size, which is on average 450 bp longer than other APOC2 cDNAs described so far. Correspondingly, the *Apoc2* gene occupies an unusually large genomic region, due to an extended 5' end. Interestingly, a variable number of tandem repeat (VNTR) in the third intron of the human APOC2 gene shows a high sequence homology and is located at the identical position in the mouse gene. Despite the high copy number of this VNTR (27 or 34 copies) only two variants were found among 11 different inbred strains. With the aid of six restriction fragment length variations in this gene cluster only two different haplotypes could be deduced, indicating that the *Apoe–c1–c2* gene cluster is highly conserved in the inbred strains that were studied. © 1993 Academic Press, Inc.

INTRODUCTION

Lipoprotein levels in plasma are important determinants of susceptibility to atherosclerosis. A large number of genes control the lipoprotein metabolism, which may lead to complex inheritance patterns in human populations. In many cases a "candidate gene" approach has been applied to reveal the defective genes in families with aberrant lipid parameters (Mahley and Rall, 1989; Goldstein and Brown, 1989). Recently, inbred mouse strains have become increasingly useful for detecting new candidate loci and dissecting polygenetic traits (Garchon *et al.*, 1991; Cornall *et al.*, 1991). Also in the field of lipoprotein metabolism new loci playing a role in the etiology of atherosclerosis have been discovered (Paigen *et al.*, 1987). In addition, the possibility of manipulating the mouse germline allows elucidation of the role of candidate genes more precisely. Transgenic mice have recently been made to carry, for instance, the apolipoprotein E (APOE) and C1 genes (Smith *et al.*, 1990; Simonet *et al.*, 1990, 1991).

The successful use of the mouse as an animal model for lipoprotein metabolism strongly depends on sufficient knowledge of the relevant mouse genes. The majority of genes known to play a role in lipoprotein metabolism have been evolutionary conserved. Most of these genes have been demonstrated at the protein level or by hybridization with heterologous DNA probes (Lusis, 1988). However, of the mouse apolipoprotein genes only *Apoe* (Reue *et al.*, 1984; Horiuchi *et al.*, 1989), *Apoa1* (Ertel Miller *et al.*, 1983), *Apoa2* (Kunisada *et al.*, 1986), and *Apoa4* (Buchberg and Kinniburgh, 1985) have been cloned so far.

In man, the APOE gene and the evolutionary related APOC1, APOC1', and APOC2 genes form a gene cluster spanning 48 kb (Davison *et al.*, 1986; Donald *et al.*, 1985; Myklebost *et al.*, 1984a; Smit *et al.*, 1988). The APOE gene is contained within a linkage group on chromosome 19 (Olaisen *et al.*, 1982) that is syntenic with a linkage group localized on mouse chromosome 7 (Saunders and Seldin, 1990).

In this study cosmid cloning was applied to demonstrate the conservation of the APOE–C1–C2 gene cluster in the mouse. Two inbred mouse strains (C57BL/Rij and 129/J) were investigated in detail and revealed two different alleles of the *Apoe–c1–c2* gene cluster. No other alleles were found among a larger set of mouse inbred strains.

MATERIALS AND METHODS

Animals Inbred mouse strains A/J, AKR, BALB/c, CBA/Ca, C3H/He, DBA/2, NZB, STR/Ort, SWR, and 129 were obtained from Harlan CPB (Zeist, The Netherlands). A C57BL/Rij mouse was derived from REPGO-TNO (Rijswijk, The Netherlands) and 129/J DNA was obtained from Dr. G. Grosveld (Erasmus University, Rotterdam).

[1] To whom correspondence should be addressed.

0888-7543/93 $5.00

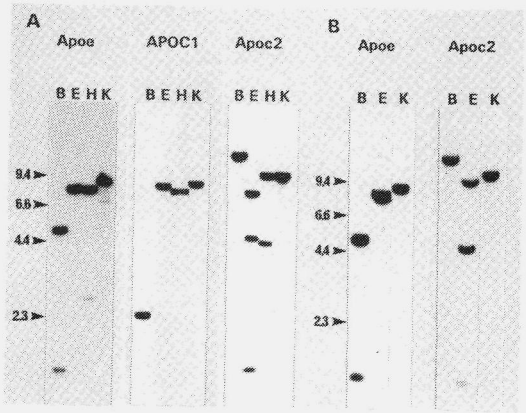

FIG. 1. Southern blot analysis of cosmid clone mAPOE27 DNA **(A)** and mouse C57BL/Rij genomic DNA **(B)** hybridized with cDNA probes of the mouse *Apoe* and the *Apoc2* genes and the human APOC1 gene. The probes used for each hybridization are indicated above each panel. DNA was digested with *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H) and *Kpn*I (K) as indicated. The molecular weight standards are shown in kilobases and are indicated at the left of each panel.

Isolation of the mouse Apoe-c1-c2 gene cluster by cosmid cloning. Two mouse cosmid libraries were constructed by cloning partial *Mbo*I digests of genomic DNA of the strains C57BL/Rij and 129/J in the *Bam*HI site of the cosmid vectors c2RB (Bates and Swift, 1983) and SuperCos1 (Stratagene), respectively. The cosmid libraries were screened by colony hybridization using a mouse *Apoe* cDNA probe, i.e., the 1.1-kb insert of pmEUC18 (Horiuchi *et al.*, 1989). The probe was labeled with [α-^{32}P]-dCTP by random priming (Multiprime labeling kit, Amersham) and hybridized (see Southern blot analysis) to the filters.

Restriction mapping. Restriction enzyme digestions were performed as recommended by the supplier (Pharmacia). The DNA fragment sizes were determined on 0.7 and 0.35% agarose gels. Restriction sites were mapped using a partial digestion protocol (Hofker *et al.*, 1986). Cosmids clones with the vector c2RB were hybridized with *Sal*I/*Bam*HI and *Sal*I/*Nru*I fragments of pBR322. SuperCos1 cosmid clones were hybridized with T3 and T7 oligonucleotides probes.

Southern blot analysis. Mouse genomic DNA was isolated from the liver (Sambrook *et al.*, 1989). Restriction enzyme digests of DNA samples were analyzed by electrophoresis through 0.7% agarose gels and blotted to Hybond N$^+$ nylon membranes (Amersham) using alkaline transfer (0.4 *M* NaOH, 1.5 *M* NaCl). Hybridizations were carried out according to Church and Gilbert (1984) and the filters were washed as recommended by the supplier. In addition to the mouse *Apoe* cDNA probe the following probes were used: a human APOC1 cDNA probe (Knott *et al.*, 1984), a human APOC2 probe (Myklebost *et al.*, 1984b), and a mouse *Apoc2* cDNA probe. The mouse *Apoc2* cDNA probe was isolated by screening a BALB/c mouse liver cDNA library (Clontech) with a human APOC2 cDNA probe, and its identity was confirmed by sequencing (Hoffer *et al.*, manuscript in preparation).

PCR amplification of the tandem repeat in the third intron of the Apoc2 gene. The cDNA sequence of the *Apoc2* gene was used to design two unique oligonucleotides flanking the minisatellite (5′ oligo: CCGATCAGCATGGATGAGAA; 3′ oligo: GCTTTTGCTGTACATGTCCC). One hundred nanograms of total mouse DNA and 100 ng of each primer were used in a 50-μl PCR reaction containing 0.2 m*M* of each dNTP, 50 m*M* KCl, 10 m*M* Tris–HCl, pH 8.4, 0.5 m*M* $MgCl_2$, 0.2 mg/ml BSA, 10% DMSO, 1 unit *Taq* polymerase (Perkin-Elmer/Cetus, Corp.) and overlaid with 50 μl of mineral oil. Reaction conditions were as follows: 1 min of denaturing at 94°C, 2 min of annealing at 57°C, and 4 min of extension at 72°C for 32 cycles in an automated PCR machine (Bio-med). Amplification products were analyzed by electrophoresis in a 2% agarose gel and visualized by staining with ethidium bromide.

RESULTS

Isolation and Characterization of the Apoe-c1-c2 Gene Cluster

A cosmid library was constructed of DNA from the mouse strain C57BL/Rij and screened with the mouse *Apoe* cDNA probe pmEUC18 (Horiuchi *et al.*, 1989). Fourteen clones were found to contain the 7.9-kb *Eco*RI fragment characteristic for *Apoe*. One of these clones, mAPOE27, was digested with four different enzymes (*Bam*HI, *Eco*RI, *Hin*dIII, and *Kpn*I) and analyzed by Southern blotting (Fig. 1A). Hybridization with pmEUC18 revealed fragments identical in size to those obtained by hybridizing this probe to mouse genomic DNA (Fig. 1B). These hybridization data confirm the presence of the *Apoe* gene in clone mAPOE27. In addition, the entire *Apoe-c1-c2* cluster is contained within cosmid mAPOE27, as can be deduced from the hybridization results with the human APOC1 (Fig. 1A) and APOC2 (data not shown) cDNA probes. The *Apoc1* gene is confined to a 2.3-kb *Bam*HI fragment, and the *Apoc2* gene to a 9.5-kb *Eco*RI fragment.

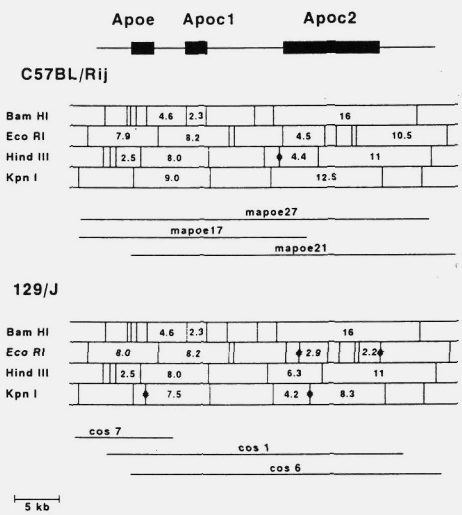

FIG. 2. Restriction map of the *Apoe-c1-c2* gene cluster and the location of the *Apoe*, *Apoc1*, and *Apoc2* genes (solid boxes). The map of the gene cluster for the strain C57BL/Rij **(A)** and that for the strain 129/J **(B)** are given for the restriction enzymes indicated at the left of both maps. The positions of the cosmids that were used for characterization are shown below each map. Restriction site variants are indicated by asterisks.

TABLE 1

Restriction Fragment Length Variations in the *Apoe–C1–C2* Gene Cluster

Strain	*Apoe* gene		*Apoc2* gene					
	*Eco*RI[a]	*Kpn*I[a]	*Eco*RI[a]	*Eco*RI[a]	*Hin*dIII[a]	*Kpn*I[a]	VNTR[b]	Haplotype[c]
A/J	7.9	9.0	4.5	10.5	4.4	12.5	1.4	1
AKR	7.9	9.0	4.5	10.5	4.4	12.5	1.4	1
BALB/c	7.9	9.0	4.5	10.5	4.4	12.5	1.4	1
CBA/Ca	7.9	9.0	4.5	10.5	4.4	12.5	1.4	1
C3H/He	8.0	7.5	1.6 + 2.9	2.2	6.3	8.3 + 4.2	1.2	2
C57BL/Rij	7.9	9.0	4.5	10.5	4.4	12.5	1.4	1
DBA/2	7.9	9.0	4.5	10.5	4.4	12.5	1.4	1
NZB	8.0	7.5	1.6 + 2.9	2.2	6.3	8.3 + 4.2	1.2	2
STR/Ort	7.9	9.0	4.5	10.5	4.4	12.5	1.4	1
SWR	7.9	9.0	4.5	10.5	4.4	12.5	1.4	1
129	8.0	7.5	1.6 + 2.9	2.2	6.3	8.3 + 4.2	1.2	2

[a] The pattern of restriction fragments found for the *Apoe* and *Apoc2* genes after digestion with the enzymes, as indicated above each column, is given in kilobases.

[b] The length of the PCR amplification products found for the VNTR in the *Apoc2* gene is given in kilobases.

[c] The haplotypes 1 and 2, correspond to the two different combinations of the restriction fragment length variations as shown in the table.

A restriction map of the *Apoe-c1-c2* locus was constructed on the basis of three overlapping cosmid clones (Fig. 2A). This map shows that the mouse *Apoc1* gene as well as the mouse *Apoc2* gene are located downstream of the *Apoe* gene at a distance of about 4.0 and 14.5 kb, respectively. The 9.5-kb *Eco*RI fragment detected with *Apoc2* maps adjacent to the vector in cosmid mAPOE27 and originates from a 10.5-kb genomic *Eco*RI fragment.

Apolipoprotein C2

The structure of the *Apoc2* gene was investigated further by isolating cDNA clones from a mouse liver cDNA library. One of these clones was found to contain a 965-bp insert that is 471 bp larger than the human cDNA (Mycklebost *et al.*, 1984b). Sequence data (EMBL Accession No. Z15090; Hoffer *et al.*, manuscript in preparation) show that the 3′ portion of the cDNA (476 bp) includes exons 2–4. This part encompasses the entire coding region of the gene and is 65.5% identical to the human cDNA. The complete 965-bp cDNA hybridizes not only to the same 10.5-kb *Eco*RI fragment as the complete human APOC2 cDNA, but also to two additional *Eco*RI fragments of 4.5 and 1.3 kb (Figs. 1A and 1B) that are located 5′ to the 10.5-kb fragment.

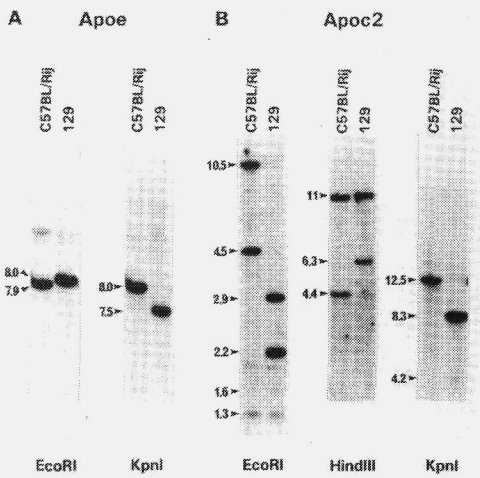

FIG. 3. Autoradiograms of Southern blots containing genomic DNA of C57BL/Rij and 129 mice digested with *Eco*RI, *Hin*dIII, or *Kpn*I. *Apoe* and *Apoc2* CDNA probes were used to reveal the RFLVs (**A** and **B**, respectively).

Genetic Variability of the Mouse Apoe-c1-c2 Gene Cluster

In addition to the cosmid clones isolated from the C57BL/Rij strain, similar cosmid clones have been isolated from the mouse strain 129/J. The restriction map of these clones reveals the same general structure of the gene cluster (Fig. 2B). However, between these two mouse strains several restriction fragment length variations (RFLVs) can be observed. The gene cluster of the 129/J mouse contains four additional restriction sites for the enzymes studied, whereas one restriction site was lost compared to C57BL/Rij. Also, an insertion of approximately 100 bp has occurred upstream of the *Apoe* gene of the 129/J mouse, giving rise to a larger *Eco*RI fragment. Table 1 shows the distribution of six of these RFLVs across 11 inbred strains. Four RFLVs were detected with the *Apoc2* cDNA probe, when using the restriction enzymes *Eco*RI, *Hin*dIII, and *Kpn*I (Fig. 3). Two RFLVs were detected with the *Apoe* cDNA probe in *Eco*RI- or *Kpn*I-digested (Lusis *et al.*, 1987) DNA.

Evolutionary Conservation of the Tandem Repeat within Intron 3 of Apoc2

Intron 3 of the human APOC2 gene contains a variable number of tandem repeat (VNTR) that is 35-bp

```
                                                                                    number
Mouse    g c t c c a g a - c c c c a g c c c - t c c t c c - - - c t c a g a c a c a g g    1
         g c t c c a g a - c c c c a g c c c - t c c t c c a c t c t c a g a c t c a g g    1
         g c t c c a g a - c c c c a g c c c - t c c t c c - - - c t c a g a c c c a g g    1
         g c t c c a g a - c c c c a g c c c - t c c t c c - - - c t c a g a c t c a g g    2
         g g t c c a g g - c c c t a t c c c - t c c t c c - - - c t c a g a c c c a g g    1
         g g t c c a g a - c c c c a g c c c - t c c t c c - - - c t c a g a c t c a g g    2
         g g t c c a g a - c c c c a g c c c - t c c t c c a c t c t c a g a c t c a g g    1
         g g t c c a g a - c c c c a g c c c - t c c t c c - - - c t c a g a c c c a g g    5
         g g t c c a g a - c c c c a g c c c - t c c t c c - - - c t c a g a c a c a g g   16

Homology * * * * * *   * * * * * * * * *  * * * * * *    * * * * * * *  * * * *

Human    a g t c c a g g n c c c c a g c c c n t c c t c c - - - c t c a g a c c c a g g
```

FIG. 4. The nine variants of the tandem repeat determined from intron 3 of the *Apoc2* gene compared with the consensus sequences found in the human gene (Das *et al.*, 1987). The number found for each variant is indicated at the right. Homologous nucleotides between man and mouse are indicated by asterisks.

long and consists of five or six copies (Das *et al.*, 1987; Hegele and Tu, 1991). This intron was partly sequenced in the mouse and a similar VNTR was found (Fig. 4). Among 30 repeat units nine different variants have been found. The most common repeat unit is 35 bp in length. In C57BL/Rij, this VNTR consists of approximately 34 copies of the repeat unit. Because only a few mismatches exist between the consensus sequences of both species, the sequence homology between mouse and man is very high (85.5%) (Fig. 4). The high copy number of this VNTR in mice prompted us to investigate its variability. Some 11 inbred mouse strains were tested by PCR for the variability of the VNTR. However, only two alleles of 1.2 and 1.4 kb in length were found, corresponding to 27 and 34 copies, respectively (Fig. 5, Table 1).

DISCUSSION

In the human genome the APOE–C1–C2 genes are contained within a gene cluster spanning 48 kb (Smit *et al.*, 1988). The APOC1 gene and the APOC2 gene are located downstream of the APOE gene at a distance of 4 and 40 kb, respectively (Davison *et al.*, 1986; Smit *et al.*, 1988), and are both transcribed in the same orientation as the APOE gene. The pseudogene of APOC1 is located between the APOC1 and APOC2 genes.

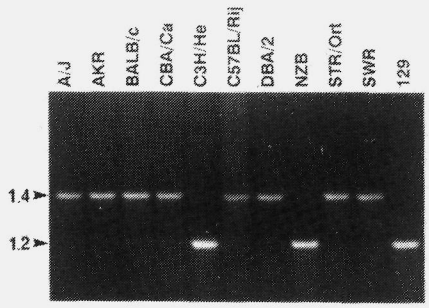

FIG. 5. PCR amplification of the VNTR of *Apoc2* in 11 different inbred mouse strains. Genomic mouse DNA was used in a PCR reaction with primers in exon 3 and 4. PCR products were analyzed by agarose gel electrophoresis.

Our hybridization data reveal that the equivalents of the human APOC1 and APOC2 genes are also present in the mouse and are located 4 and 15 kb downstream of the *Apoe* gene, respectively. The isolation of cDNA clones for *Apoc1* (Simonet *et al.*, 1991) and *Apoc2* (this paper) indicates that the genes are both actively transcribed. Interestingly, the *Apoc2* cDNA described here has an extended 5′ end, and these additional sequences can be detected in the cosmid 5′ to exon 2–4 of the *Apoc2* gene, suggesting a much more complicated promoter region than described so far.

The recent duplication of the human APOC1 gene, resulting in the APOC1′, is absent in the mouse. This finding confirms a recent hypothesis by Raisonnier (1991), who studied the sequence divergence of the APOC1 and the APOC1′ genes in man. Based on calculations of the rate of nucleotide divergence, it was estimated that the human APOC1 duplication took place after the divergence of rodents and primates. Coordinate expression could be an important factor leading to the conservation of the structure of this locus. One DNA element located just 5′ of APOC1′ determines the liver-specific expression of both APOE and APOC1 (Simonet *et al.*, 1991). The mouse cosmid clones described here allow us to search for this DNA element on the basis of sequence homology and function.

Genetic Variation

A large fraction of the human VNTRs cross-hybridize with mouse DNA, indicating that sequence conservation between mouse and humans is common (Julier *et al.*, 1990). One such example of evolutionary conservation is presented here. The VNTR of intron 3 of the human APOC2 gene has a high sequence homology (up to 85%)

between the individual repeat units and hybridizes also on mouse genomic DNA (Das *et al.*, 1987). Interestingly, this VNTR could also be found in the mouse gene and the position of the VNTR has been conserved as well. If this positional conservation were a more general characteristic of VNTRs, these markers would be very helpful —in addition to genes—for delineating the synteny between genetic maps of different organisms.

Despite the high copy number of this repeat in the mouse (27 and 34 copies), only two size variants were found among 11 inbred strains. Since these inbred strains were chosen on the basis of a high degree of variability of many other loci (Atchley and Fitch, 1991), the low number of variants detected with this VNTR may be due to the low polymorphism content of the VNTR itself. This finding may reflect a low degree of allelic variation of the APOE–C1–C2 gene cluster. The latter hypothesis is supported by the detection of only two different haplotypes for the *Apoe–c1–c2* gene cluster (Table 1) when six additional variant sites were examined. The absence of genetic variation of this gene cluster could be a result of selection during the generation of inbred mouse strains. Investigation of the number of alleles at certain key loci, such as the APOE–C1–C2 locus, augments other mapping strategies that are being used to unravel the genetic factors controlling lipoprotein levels in mice.

ACKNOWLEDGMENTS

The authors are grateful to Cisca Wijmenga for her helpful discussions and Dr. G Grosveld of the department of Cell Biology, Erasmus University, for his generous support. M.J.V H. is supported by HGO-TNO (900-539-052). M.H.H. receives a fellowship of the Royal Netherlands Academy of Arts and Science.

REFERENCES

Atchley, W R., and Fitch, W. M. (1991). Gene trees and the origins of inbred strains of mice. *Science* **254:** 554–558.

Bates, F. P., and Swift, R. A. (1983). Double *cos* site vectors: Simplified cosmid cloning. *Gene* **26:** 137–146.

Buchberg, A. M., and Kinniburgh, A J. (1985). Induction of liver apolipoprotein A-IV mRNA in porphyric mice. *Nucleic Acids Res* **13:** 1953–1963.

Church, G M., and Gilbert, W. (1984). Genomic sequencing *Proc Natl Acad. Sci USA* **81:** 1991–1995.

Cornall, R. J., Prins, J-B., Todd, J. A., Pressey, A , DeLarato, N. H., Wicker, L. S., and Peterson, L. B. (1991) Type I diabetes in mice is linked to the interleukin-1 receptor and *Lsh/Ity/Bcg* genes on chromosome 1. *Nature* **353:** 262–265.

Das, H. K., Jackson, C L., Miller, D A., Leff, T., and Breslow, J. L. (1987). The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron. *J Biol. Chem* **262:** 4787–4793.

Davison, P. J., Norton, P., Wallis, S. C , Gill, I., Cook, M., Williamson, R., and Humphries, S. E (1986). There are two gene sequences for human apolipoprotein C1 (APO C1) on chromosome 19, one of which is 4 kb from the gene for apo E. *Biochem Biophys Res Commun.* **136:** 876–884.

Donald, J. A., Wallis, S. C., Kessling, A., Tippett, P., Robson, E B., Ball, S , Davies, K. E., Scambler, P., Berg, K., Heiberg, A., Williamson, R., and Humphries, S. E. (1985). Linkage relationships of the gene for apolipoprotein CII with loci on chromosome 19. *Hum Genet.* **69:** 39–43.

Ertel Miller, J. C., Barth, R. K., Shaw, P. H., Elliott, R W , and Hastie, N. D. (1983). Identification of a cDNA clone for mouse apoprotein A-I (apo A-I) and its use in characterization of apo A-I expression in liver and small intestine. *Proc. Natl. Acad. Sci. USA* **80:** 1511–1515.

Garchon, H-J., Bedossa, P., Eloy, L., and Bach, J-F. (1991). Identification and mapping to chromosome 1 of a susceptibility locus for periinsulitis in non-obese diabetic mice. *Nature* **353:** 260–262.

Goldstein, J. L., and Brown, M. S. (1989). Familial hypercholesterolemia. *In* "The Metabolic Basis of Inherited Disease" (C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle, Eds.) 6th ed., pp. 1214–1250, McGraw–Hill, New York.

Hegele, R. A., and Tu, L. (1991). Variation within intron 3 of the apolipoprotein CII gene. *Nucleic Acid Res.* **19:** 3162.

Hofker, M. H., van Ommen, G. J. B., Bakker, E., Burmeister, M , and Pearson, P. L. (1986). Development of additional RFLP probes near the locus for Duchene muscular dystrophy by cosmid cloning of the DXS84 (754) locus. *Hum Genet.* **74:** 275–279.

Horiuchi, K , Tajima, S., Menju, M., and Yamamoto, A. (1989). Structure and expression of mouse apolipoprotein E gene. *J Biochem.* **106:** 98–103.

Julier, C., De Gouyon, B., Georges, M., Guénet, J-L., Nakamura, Y., Avner, P , and Lathrop, G. M. (1990). Minisatellite linkage maps in the mouse by cross-hybridization with human probes containing tandem repeats. *Proc. Natl. Acad Sci. USA* **87:** 4585–4589.

Knott, T. J., Robertson, L. M., Priestly, L. M., Urdea, M., Wallis, S., and Scott, J. (1984). Characterization of mRNAs encoding the precursor for human apolipoprotein CI. *Nucleic Acids Res.* **12:** 3909–3915.

Kunisada, T., Higuchi, K., Aota, S-I , Takeda, T., and Yamagishi, H (1986). Molecular cloning and nucleotide sequence of cDNA for murine senile amyloid protein: Nucleotide substitutions found in apolipoprotein A-II cDNA of senescence accelerated mouse (SAM). *Nucleic Acid Res* **14:** 5729–5740.

Lusis, A. J., Taylor, B. A., Quon, D., Zollman, S., and LeBoeuf, R. C (1987). Genetic factors controlling structure and expression of apolipoprotein B and E in mice. *J Biol. Chem.* **262:** 7594–7604

Lusis, A. J. (1988). Genetic factors affecting blood lipoproteins: The candidate gene approach. *J Lipid Res* **29:** 397–429.

Mahley, R. W , and Rall, S. C., Jr. (1989) Type III hyperlipoproteinemia (Dysbetalipoproteinemia): The role of apolipoprotein E in normal and abnormal lipoprotein metabolism. In "The Metabolic Basis of Inherited Disease" (C. R Scriver, A L. Beaudet, W. S. Sly, and D. Valle, Eds.), 6th ed., pp. 1195–1213, McGraw–Hill, New York.

Mycklebost, O., Rogne, S., Olaisen, B., Gedde-Dahl, T., Jr., and Prydz, H. (1984a). The locus for apolipoprotein CII is closely linked to the apolipoprotein E locus on chromosome 19 in man. *Hum. Genet.* **67:** 309–312.

Myklebost, O , Williamson, B., Markham, A. F., Myklebost, S. R., Rogers, J., Woods, D E., and Humphries, S. E (1984b). The isolation and characterization of cDNA clones for human apolipoprotein CII *J Biol Chem* **259:** 4401–4404.

Paigen, B., Mitchell, D., Reue, K., Morrow, A., and Lusis, A. J. (1987) *Ath-1*, a gene determining atherosclerosis susceptibility and high density lipoprotein levels in mice. *Proc Natl. Acad. Sci USA* **84:** 3763–3767.

Olaisen, B., Teisberg, P., and Gedde-Dahl, T., Jr. (1982). The locus for apolipoprotein E (apo E) is linked to the complement component C3 (C3) locus in chromosome 19 in man. *Hum Genet.* **62:** 233–236

Raisonnier, A. (1991). Duplication of the apolipoprotein C-I gene occurred about forty million years ago. *J Mol Evol* **32:** 211–219.

Reue, K. L., Quon, D H., O'Donnell, K. A., Dizikes, G J , Fareed, G. C , and Lusis, A. J. (1984). Cloning and regulation of messenger RNA for mouse apolipoprotein E *J Biol Chem* **259:** 2100–2107

Sambrook, J , Fritsch, E. F , and Maniatis, T. (1989) "Molecular Cloning: A Laboratory Manual." 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Saunders, A. M , and Seldin, M. F. (1990). A molecular genetic linkage map of mouse chromosome 7. *Genomics* **8:** 525–535.

Simonet, W. S., Bucay, N., Lauer, S. J., Wirak, D. O., Stevens, M. E., Weisgraber, K. H., Pitas, R. E., and Taylor, J. M. (1990). In the absence of a downstream element, the apolipoprotein E gene is expressed at high levels in kidneys of transgenic mice. *J. Biol. Chem.* **265:** 10809–10812.

Simonet, W. S., Bucay, N., Pitas, R. E., Lauer, S. J., and Taylor, J. M. (1991). Multiple tissue-specific elements control the apolipoprotein E/C-I gene locus in transgenic mice. *J. Biol. Chem.* **266:** 8651–8654.

Smit, M., van der Kooy-Meijs, E., Frants, R. R., Havekes, L., and Klasen, E. C. (1988). Apolipoprotein gene cluster on chromosome 19: Definite localization of the APOC2 gene and the polymorphic *HpaI* site associated with type III hyperlipoproteinemia. *Hum. Genet.* **78:** 90–93.

Smith, J. D., Plump, A. S., Hayek, T., Walsh, A., and Breslow, J. L. (1990). Accumulation of human apolipoprotein E in the plasma of transgenic mice. *J. Biol. Chem.* **265:** 14709–14712.

CHAPTER 3

THE MOUSE APOLIPOPROTEIN C1 GENE: STRUCTURE AND EXPRESSION.

M.J.V. Hoffer, M.M. van Eck, L.M. Havekes, M.H. Hofker, and R.R. Frants.

Genomics: in press

The Mouse Apolipoprotein C1 Gene: Structure and Expression

MARIËTTE J. V. HOFFER, MARGA M. VAN ECK, LOUIS M. HAVEKES,* MARTEN H. HOFKER,[1] AND RUNE R. FRANTS

*MGC-Department of Human Genetics, Leiden University, P.O. Box 9503, 2300 RA Leiden, The Netherlands; and *IVVO-TNO, Gaubius Laboratory, P.O. Box 430, 2300 AK Leiden, The Netherlands*

Received May 18, 1993

We have isolated and characterized cDNA and genomic clones containing the mouse apolipoprotein C1 (*Apoc1*) gene. The *Apoc1* gene is part of the *Apoe–c1–c2* gene cluster and is located 3.4 kb 3′ of the *Apoe* gene. The mouse *Apoc1* gene spans a region of approximately 3.3 kb and consists of four exons. The exon–intron structure is similar to those of human and baboon genes, although in mouse introns 2 and 3 are smaller. Significant sequence homology is found between man and mouse in the promoter and exonic regions (80 and 67%, respectively). Northern blotting and primer extension analysis of mouse RNA showed that a major transcript 409 bp in size is expressed primarily in fetal and adult liver. The mouse *Apoc1* cDNA contains an open reading frame encoding a protein of 88 amino acids, including a signal peptide of 26 amino acid residues. Comparisons of the deduced amino acid sequence of mouse apoC1 with the human, baboon, rat, and dog sequences showed discrete regions with a high degree of conservation. The delineation of the sequence and structural organization of the mouse *Apoc1* gene is an essential step in enhancing the use of mouse models to study the function of apoC1 in the lipoprotein metabolism. © 1993 Academic Press, Inc.

INTRODUCTION

Human apolipoprotein C1 (apoC1) is a 57-amino-acid peptide that is primarily produced in the liver (Knott *et al.*, 1984) and is associated with chylomicron, very-low-density lipoprotein, and high-density lipoprotein (Breslow, 1988). The function of apoC1 in the lipoprotein metabolism remains unclear. ApoC1 is known to activate lecithin:cholesterol acyltransferase (LCAT) *in vitro* (Soutar *et al.*, 1975). In addition, apoC1 can inhibit the binding of apoE to the low-density lipoprotein receptor-related protein (Weisgraber *et al.*, 1990). More recent studies showed that this inhibition can also be found for the low-density lipoprotein receptor (Sehayek and Eisenberg, 1991). These studies could be in agreement with the data obtained with a human APOC1 transgenic mouse model in which overexpression of human apoC1 resulted in an increase of the plasma triglyceride levels (Simonet *et al.*, 1991). Apparently, apoC1 plays a role in regulating the catabolism of triglyceride-rich lipoproteins.

Due to the advances in transgenic technology, mice have become increasingly useful for investigation of lipoprotein metabolism. The correct interpretation of results obtained with transgenic mice strongly depends on sufficient knowledge of the genes involved in the mouse lipoprotein metabolism. The human APOC1 gene is part of a gene cluster that is located on the long arm of chromosome 19 and includes the genes for apoE, apoC1, pseudo apoC1, and apoC2 (Humphries *et al.*, 1984; Davison *et al.*, 1986; Smit *et al.*, 1988). We have demonstrated a similar gene organization in mouse with the exception that the mouse gene cluster does not contain a pseudo *Apoc1* gene (Hoffer *et al.*, 1993a). The genomic structure and sequences of the *Apoe* (Reue *et al.*, 1984; Horiuchi *et al.*, 1989) and the *Apoc2* genes (Hoffer *et al.*, 1993b) have been determined previously. This paper describes the characterization of the mouse *Apoc1* gene. We have determined the mouse *Apoc1* cDNA sequence and the exon–intron structure and analyzed the expression of the gene.

MATERIALS AND METHODS

Isolation and sequencing of cDNA clones. A λgt11 cDNA library derived from liver mRNA of an adult male BALB/c mouse (Clontech) was screened with a *DdeI* fragment of human APOC1 cDNA (Knott *et al.*, 1984) containing exons 1–4. Inserts of positive clones were isolated by PCR amplification using primers (a. 5′-GGTGGCGACGACTCCTGGAG-3′; b. 5′-GGTAATGGTAGCGACCGGCG-3′) flanking the *EcoRI* cloning site. A total of 30 μl of phage lysate and 100 ng of each primer were used in a 50-μl PCR reaction containing 0.2 m*M* of each dNTP, 50 m*M* KCl, 10 m*M* Tris–HCl (pH 8.4), 2.5 m*M* $MgCl_2$, 0.2 mg/ml BSA, 10% DMSO, 1 unit *Taq* polymerase (Perkin–Elmer–Cetus) and overlaid with 50 μl of mineral oil. Reaction conditions were as follows: a denaturing step of 1 min at 94°C, followed by an annealing step of 2 min at 60°C and an extension step of 4 min at 72°C for 32 cycles. PCR products were subcloned into the pT7Blue (R) T-vector (Novagen) and sequenced using the T7 sequencing kit (Pharmacia). Individual clones were sequenced in both directions using the universal primers or specific synthetic oligonucleotides.

Determination of exon–intron structure. Exons were located by hybridization with a mouse *Apoc1* cDNA probe on restriction enzyme

[1] To whom correspondence should be addressed. Telephone: 31-71-276099. Fax: 31-71-276075.

0888-7543/93 $5.00

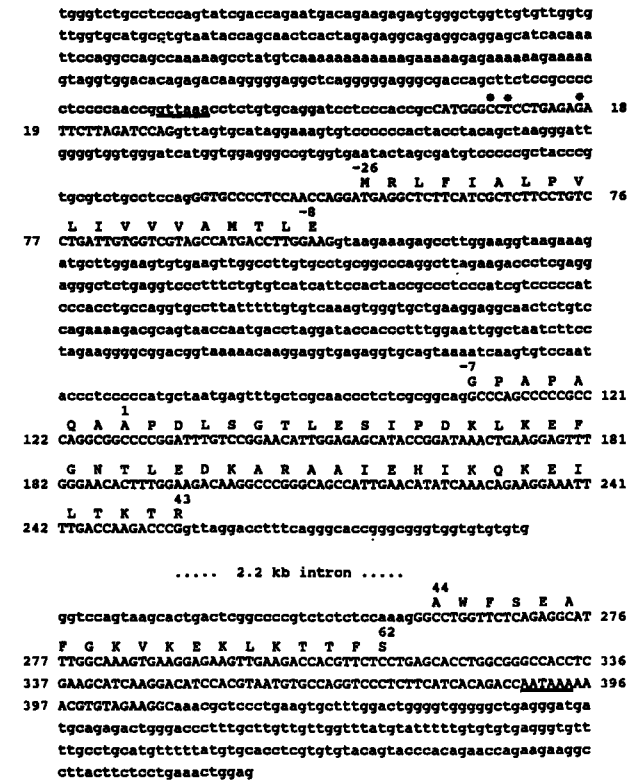

FIG. 1. Nucleotide and deduced amino acid sequences of mouse *Apoc1* gene. Exon nucleotide sequences and the flanking intron sequences are shown in uppercase and lowercase, respectively. The transcription start site is marked +1 and only the exonic sequences are numbered. The deduced amino acid sequence, in one-letter code, is indicated above the nucleotide sequence; the predicted signal sequence is indicated by negative numbers. Asterisks above the sequence indicate the first nucleotide of the cDNA clones mAPOC1c9, c8, and c16, respectively. The putative TATA box and polyadenylation signal are double underlined. (EMBL Accession Nos. Z22660 and Z22661.)

fragments from cosmid mAPOE27 (Hoffer *et al.*, 1993a). For sequencing analysis, positive DNA fragments were subcloned into the pUC19 or M13mp18/mp19. All exons, exon–intron borders, and 5'- and 3'-flanking regions of *Apoc1* were sequenced using the T7 sequencing kit (Pharmacia) and universal primers or specific synthetic oligonucleotides.

Primer extension analysis. Poly(A)$^+$ RNA was isolated using the polyATract mRNA isolation kit (Promega). A synthetic oligonucleotide (5'-CTTCCAAGGTCATGGCTACG-3') was end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase and annealed to 2 μg poly(A)$^+$ RNA isolated from liver. The primer was extended using AMV reverse transcriptase (Promega) as described by Geliebter *et al.* (1986). The extension product was run through a 6% polyacrylamide sequencing gel. An M13mp18 sequencing ladder served as a size standard to determine the length of the extension product.

mRNA isolation and characterization by Northern blot analysis. Total cellular RNA from several tissues, including liver from a 12-day-old fetus, was isolated according to Chomczynski and Sacchi (1987). Peritoneal macrophages were isolated after induction by an intraperitoneal injection of 0.5 ml sterile 0.8% glycogen. After 1 day, a booster was administered and 3 h later the cells were collected. RNA samples were subjected to electrophoresis on a 1.2% (w/v) denaturing agarose gel containing 7.5% formaldehyde and then transferred to a nylon membrane (Hybond N$^+$; Amersham) according the protocol provided by the manufacturer. Hybridizations were performed at 55°C in a hybridization solution containing 50% formamide (Krumlauf, 1991). The cDNA probes were labeled using the random primer method (Multiprime labeling kit; Amersham) with [α-^{32}P]dCTP. The filters were washed twice in 2× SSC (SSC: 0.15 *M* NaCl, 15 m*M* sodium citrate), 0.5% SDS for 15 min at 65°C and twice in 1× SSC, 0.5% SDS for 15 min.

RESULTS

Isolation and Sequence Analysis of Mouse Apoc1 cDNA Clones

A liver cDNA library was screened with a human APOC1 cDNA probe (Knott *et al.*, 1984) and yielded

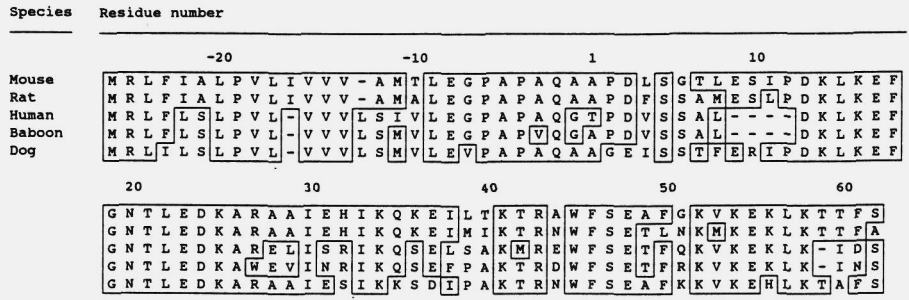

FIG. 2. Comparison of the deduced amino acid sequences of the mouse, human, baboon, rat, and dog apoC1. Negative numbers are used for amino acid residues in the signal peptides. Deletions are indicated by dashes. Boxes indicate homologies to the mouse sequence. Sequence data for rat apoC1 are from Shen and Howlett (1989); for human apoC1, from Knott *et al.* (1984); for baboon apoC1, from Pastorcic *et al.* (1992); and for dog apoC1, from Luo *et al.* (1989).

three putative mouse *Apoc1* clones. A large clone, mAPOC1c16, was subcloned and completely sequenced. The sequence of this clone was determined on both strands (Fig. 1). The 3′-end of this clone was followed by a poly(A)$^+$ tail. The three different cDNA clones map 11 bp from each other at the 5′-end. However, the 5′-end of clone mAPOC1c16 was unexpectedly large. Sequence analysis revealed that a part of these additional sequences was homologous with genomic sequences corresponding to intron 1. This finding was confirmed by sequence analysis of the two other cDNA clones, mAPOC1c8 and mAPOC1c9, which were properly spliced and contained only sequences corresponding to exons 1 and 2. Comparison of mouse *Apoc1* with human APOC1 cDNA sequences revealed 63% homology. The cDNA has an open reading frame encoding a protein of 88 amino acid residues. Figure 2 shows the deduced amino acid sequence of mouse apoC1 and its alignment with all other known apoC1 sequences. The amino acid sequence is 5 residues longer than that of the corresponding human protein (Knott *et al.*, 1984). When the mouse apoC1 is compared to the human apoC1 amino acid sequence, the additional residues in the mouse protein occur at position 9–12 (4 residues) and at position 59 (1 residue).

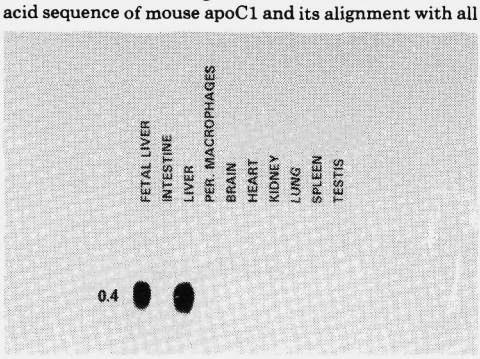

FIG. 3. Northern blot analysis. Total cellular RNA (10 μg) isolated from several tissues, as indicated above each lane, was examined by Northern blot analysis. The Northern blot was probed with a fragment of mouse *Apoc1* cDNA containing exons 2–4. The length of the mRNA in kilobases is indicated at the left. The autoradiogram was exposed overnight.

Expression of Apoc1

Northern blot analysis using a cDNA probe containing exons 2–4 of *Apoc1* showed an mRNA of approximately 400 bp in size, in adult and fetal liver only (Fig. 3). No signal was detected in other tissues, including peritoneal macrophages, lung, and testis. A control

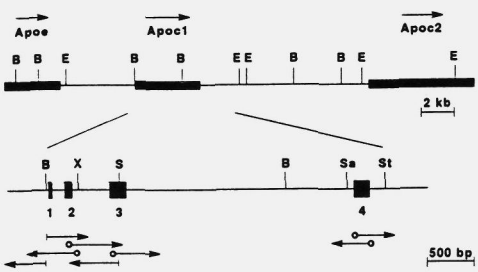

FIG. 4. Restriction map for the mouse *Apoc1* gene. At the top, a map of the mouse gene cluster containing the *Apoe*, the *Apoc1*, and the *Apoc2* genes is shown. The genes are indicated by filled boxes and the transcriptional orientation is shown by arrows above the map. Restriction enzymes are indicated as follows: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; K, *Kpn*I. Below the cosmid map, an expanded map of the mouse *Apoc1* gene shows the exon–intron structure. Exons are indicated by filled boxes. Sequencing strategy is shown at the bottom of the figure. Arrows indicate the direction of sequencing. The nucleotide sequence was determined with specific oligonucleotides (open circles) or with universal primers (vertical bars). A partial restriction map is shown for *Sac*I (Sa) and *Stu*I (St). Other restriction enzymes are X, *Xho*I; S, *Sma*I.

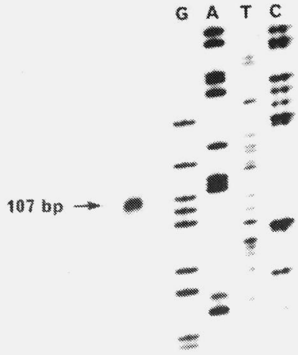

FIG. 5. Primer extension analysis of mouse *Apoc1* mRNA. Mouse poly(A)$^+$ liver RNA (1 μg) was annealed to a γ-^{32}P-labeled oligonucleotide specific for the 3′ end of exon 2 of *Apoc1*. The oligonucleotide was extended and analyzed as described under Materials and Methods. A sequencing reaction was used in parallel lanes as a size marker.

probe was used to demonstrate that the mRNA from these tissues was not degraded (data not shown).

Genomic Structure of the Mouse Apoc1 Gene

An 8.2-kb *Eco*RI fragment from cosmid mAPOE27, containing the *Apoe-c1-c2* gene cluster (Hoffer *et al.*, 1993a), was subcloned and mapped for several restriction enzymes (Fig. 4). The exon–intron structure of *Apoc1* (Fig. 4) was determined by restriction mapping and Southern blot analysis followed by sequencing of the exon–intron boundaries. The *Apoc1* gene is located 3.4 kb 3′ of the *Apoe* gene and both genes are transcribed in the same orientation. The transcriptional start site of mouse *Apoc1* was detected by primer extension analysis with mouse liver poly(A)$^+$ RNA using a primer in exon 2. Figure 5 shows an extension product after electrophoresis and demonstrates that the mouse *Apoc1* transcription start site is located 26 bp 3′ of the TATA-like sequence (GTTAAA) (Fig. 1). Sequence analysis of genomic clones and comparison with the cDNA sequence revealed that the mouse *Apoc1* gene consists of four exons and three introns spanning a region of approximately 3.3 kb (Figs. 1 and 4). Intron 1 (122 bp) starts at nucleotide 31 in the 5′-untranslated region. Intron 2 (354 bp) starts at the first nucleotide of codon −8 and is smaller than the human gene (1244 bp) (Lauer *et al.*, 1988). Intron 3 (2.2 kb) starts at the first nucleotide of codon 43. All the exon–intron junctions in the mouse *Apoc1* gene follow the GT/AG rule (Breathnach *et al.*, 1978). Hybridization with a mouse *Apoc1* cDNA probe containing exons 2–4 to mouse genomic DNA digested with several restriction enzymes showed that the mouse *Apoe-c1-c2* gene cluster contains only one copy of the *Apoc1* gene (data not shown), which is in agreement with earlier results obtained with cosmid DNA (Hoffer *et al.*, 1993a).

DISCUSSION

We have characterized and sequenced the gene for apoC1 (*Apoc1*) in mouse. Translation of the *Apoc1* cDNA predicted a protein of 88 amino acid residues, including a signal sequence of 26 amino acids on the basis of homology with sequences found for the human apoC1 protein (Knott *et al.*, 1984). This compares to 83 amino acid residues found in human and baboon apoC1 (Knott *et al.*, 1984; Pastorcic *et al.*, 1992) and 88 amino acids found in both rat and dog (Shen and Howlett, 1989; Luo *et al.*, 1989). As shown in Fig. 2, the overall amino acid sequence is approximately 72% conserved among the five species. A very strong conservation is found between amino acid residues 13 and 26. Studies with synthetic peptides of apoC1 indicated that this part of the protein is involved in activation of LCAT and that residues 32–57 represent one of the major phospholipid-binding regions of apoC1 (Soutar *et al.*, 1978).

Primer extension analysis showed that the transcription start site in the liver is located 26 bp 3′ of the TATA

```
MOUSE     aaaaaaaaaagagaaaaaagaaaaa-gtaggtggacaca--gagaca-------aggggg   -79
          :    : :  :  : :  ::::::: :: :::::::: :  ::: ::       :::::
HUMAN     acgggacaggggcagaggagaaaaacgtgggtggacagagggaggcaggcggtcagggga   -79

                                                 GC box       TATA box
                                             ************** *
MOUSE     aggctcaggggagggcgaccagc-------ttctccgccccctccccaaccggttaaa-    -26
          ::::::::: :::::: :: :: :       : : :::::::::::::: :: : ::::
HUMAN     aggctcaggaggagggagatcaacatcaacctgccccgccccctccccagcctgataaag   -19

MOUSE     --cctctgtgcagga----tcctcccaccgcCATGGGCCTCCTGAGAGATCCTTAGATCC    29
            :::  : ::::::     :::::::   :::  : ::::: :  ::       : :
HUMAN     gtcctgcgggcaggacagGACCTCCCA--ACCA-AGCCCTCCAGCAAGGATTCAGgttgg    39

MOUSE     AGgttagtgcataggaaagtgtccc-cccactacctac----agct
           : : ::::: : :::  :   ::: :: ::   :: :    : ::
HUMAN     tgctgagtgcctgggagggacacccgcctacactctgcaagaaact
```

FIG. 6. Comparison of sequences of the promoter region and exon 1 of the mouse and human APOC1 genes. Identical nucleotides are marked (:). Gaps (−) have been inserted to achieve maximum homology. Exon 1 is indicated in uppercase. The transcription start site is marked +1. The TATA and GC boxes are indicated above the sequence. Asterisks indicate homology with element I in the promoter region of human APOE (Chang *et al.*, 1990). Human data are according to Lauer *et al.* (1988).

box. The 5′ sequences of the three cDNA clones, containing exons 1–4, are found just 3′ of this start site. In addition, the cDNA library yielded a clone, mAPOC1c16, containing sequences corresponding to intron 1 that is probably due to incomplete splicing. A similar phenomenon seems to be the case for the rat APOC1 cDNA sequence (Shen and Howlett, 1989) the first 66 bp of which show a high homology (80%) with 3′ sequences of intron 1 instead of exon 1 of the mouse *Apoc1* gene, described in this paper.

The exon–intron structure resembles the overall gene structure of the human APOC1 gene with the exception that introns 2 and 3 of the mouse gene are relatively small. This can be due to the absence of *Alu*-like repeated elements that make up some of the intron sequences of the human gene (Lauer *et al.*, 1988).

The nucleotide sequence of the promoter region of the mouse gene was compared to that of the human APOC1 gene (Lauer *et al.*, 1988) (Fig. 6), and their nucleotide sequences show a homology of 80% between nucleotides −114 and −1. However, the region of high homology is limited due to the presence of a SINE repeat in the mouse promoter region located approximately 100 bp upstream of the transcription start site. Both genes contain an imperfect TATA box. Interestingly, a 15-bp sequence motif that shows 93% homology with a GC box (element I) in the human APOE gene was found (Smith *et al.*, 1988). Studies involving the APOE gene showed that element I binds Sp1 and is involved in positive regulation (Chang *et al.*, 1990). The high conservation of the sequences immediately 5′ of the TATA box suggests functional significance for gene regulation. Further studies of these putative regulatory elements will be necessary to prove any functional significance.

In most species, APOC1 is expressed primarily in the liver. However, in human, APOC1 can also be detected in monocytes differentiating into macrophages (Lauer *et al.*, 1988). In the mouse, both the *Apoe* and *Apoc2* genes are expressed in peritoneal macrophages (Basu *et al.*, 1981; Hoffer *et al.*, 1993b) but expression of *Apoc1* could not be detected in this cell type. Other studies showed limited amounts of *Apoc1* mRNA expression in skin, testis, lung, and submaxillary when a more sensitive RNase protection assay was used (Simonet *et al.*, 1991). Transgenic mice can be useful for further studies on the expression of the *Apoc1* gene and the interaction of regulatory elements in the *Apoe–c1–c2* gene cluster. The isolation and characterization of the mouse *Apoc1* gene are essential steps for generating an *Apoc1*-deficient mouse model that could be useful for studying the function of apoC1 *in vivo*.

ACKNOWLEDGMENTS

The authors acknowledge Dr. K. Willems van Dijk for his helpful discussions. This work was supported by HGO-TNO (900-539-052) and the Royal Netherlands Academy of Arts and Science (M.H.H.)

REFERENCES

Basu, S. K., Brown, M. S., Ho, Y. K., Havel, R. J., and Goldstein, J. L. (1981). Mouse macrophages synthesize and secrete a protein resembling apolipoprotein E. *Proc. Natl. Acad. Sci. USA* **78:** 7545–7549.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon–intron boundaries. *Proc. Natl. Acad. Sci. USA* **75:** 4853–4857.

Breslow, J. L. (1988). Apolipoprotein genetic variation and human disease. *Physiol. Rev.* **68:** 85–132.

Chang, D. J., Paik, Y-K., Leren, T. P., Walker, D. W., Howlett, G. J., and Taylor, J. M. (1990). Characterization of a human apolipoprotein E gene enhancer element and its associated protein factors. *J. Biol. Chem.* **265:** 9496–9504.

Chomczynski, P., and Sacchi, N. (1987). Single step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction. *Anal. Biochem.* **162:** 156–159.

Davison, P. J., Norton, P., Wallis, S. C., Gill, L., Cook, M., Williamson, R., and Humphries, S. E. (1986). There are two gene sequences for human apolipoprotein CI (APO CI) on chromosome 19, one of which is 4 kb from the gene for APO E. *Biochem. Biophys. Res. Commun.* **136:** 876–884.

Geliebter, J., Zeff, R. A., Melvold, R. W., and Nathenson, S. G. (1986). Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequencing analysis: K^{bm3} and K^{bm8}. *Proc. Natl. Acad. Sci. USA* **83:** 3371–3375.

Hoffer, M. J. V., Hofker, M. H., van Eck, M. M., Havekes, L. M., and Frants, R. R. (1993a). Evolutionary conservation of the mouse apolipoprotein *e-c1-c2* gene cluster: Structure and genetic variability in inbred mice. *Genomics* **15:** 62–67.

Hoffer, M. J. V., van Eck, M. M., Havekes, L. M., Hofker, M. H., and Frants, R. R. (1993b). Structure and expression of the mouse apolipoprotein C2 gene. *Genomics* **17:** 45–51.

Horiuchi, K., Tajima, S., Menju, M., and Yamamoto, A. (1989). Structure and expression of mouse apolipoprotein E gene. *J. Biochem.* **106:** 98–103.

Humphries, S. E., Berg, K., Gill, L., Cumming, A. M., Robertson, F. W., Stalenhoef, A. F. H., Williamson, R., and Borresen, A. L. (1984). The gene for apolipoprotein CII is closely linked to the gene for apolipoprotein E on chromosome 19. *Clin. Genet.* **26:** 389–396.

Knott, T. J., Robertson, L. M., Priestly, L. M., Urdea, M., Wallis, S., and Scott, J. (1984). Characterization of mRNAs encoding the precursor for human apolipoprotein CI. *Nucleic Acids Res.* **12:** 3909–3915.

Krumlauf, R. (1991). Northern blot analysis of gene expression. *In* "Gene Transfer and Expression Protocols" (E. J. Murray, Ed.), pp. 307–360, Humana Press, Clifton, NJ.

Lauer, S. J., Walker, D., Elshourbagy, N. A., Reardon, C. A., Levy-Wilson, B., and Taylor, J. M. (1988). Two copies of the human apolipoprotein C-I gene are linked closely to the apolipoprotein E gene. *J. Biol. Chem.* **263:** 7277–7286.

Luo, C-C., Li, W-H., and Chan, L. (1989). Structure and expression of dog apolipoprotein A-1, E, and C-1 mRNAs: Implications for the evolution and functional constraints of apolipoprotein structure. *J. Lipid Res.* **30:** 1735–1746.

Pastorcic, M., Birnbaum, S., and Hixson, J. E. (1992). Baboon apolipoprotein C-I: cDNA and gene structure and evolution. *Genomics* **13:** 368–374.

Reue, K. L., Quon, D. H., O'Donnell, K. A., Dizikes, G. J., Fareed, G. C., and Lusis, A. J. (1984). Cloning and regulation of messenger RNA for mouse apolipoprotein E. *J. Biol. Chem.* **259:** 2100–2107.

Sehayek, E., and Eisenberg, S. (1991). Mechanisms of inhibition by apolipoprotein C of apolipoprotein E-dependent cellular metabo-

lism of human triglyceride-rich lipoproteins through the low density lipoprotein receptor pathway. *J. Biol. Chem.* **266:** 18259–18267.

Shen, P. Y., and Howlett, G. J. (1989). Nucleotide sequence for rat apolipoprotein C-I. *Nucleic Acids Res.* **17:** 6405.

Simonet, W. S., Bucay, N., Pitas, R. E., Lauer, S. J., and Taylor, J. M. (1991). Multiple tissue-specific elements control the apolipoprotein E/C-I gene locus in transgenic mice. *J. Biol. Chem.* **266:** 8651–8654.

Smit, M., van der Kooy-Meya, E., Franta, R. R., Havekes, L. M., and Klasen, E. C. (1988). Apolipoprotein gene cluster on chromosome 19. *Hum. Genet.* **78:** 90–93.

Smith, J. D., Melian, A., Leff, T., and Breslow, J. L. (1988). Expression of the human apolipoprotein E gene is regulated by multiple positive and negative elements. *J. Biol. Chem.* **263:** 8300–8303.

Soutar, A. K., Garner, C. W., Baker, H. N., Sparrow, J. T., Jackson, R. L., Gotto, A. M., and Smith, L. C. (1975). Effect of the human plasma apolipoproteins and phosphatidylcholine acyl donor on the activity of lecithin:cholesterol acyltransferase. *Biochemistry* **14:** 3057–3064.

Soutar, A. K., Sigler, G. F., Smith, L. C., Gotto, A. M., Jr., and Sparrow, J. T. (1978). Lecithin:cholesterol acyltransferase activation and lipid binding by synthetic fragments of apolipoprotein C-I. *Scand. J. Clin. Lab. Invest.* **38** (Suppl.): 53–58.

Weisgraber, K. H., Mahley, R. W., Kowal, R. C., Herz, J., Goldstein, J. L., and Brown, M. S. (1990). Apolipoprotein C-I modulates the interaction of apolipoprotein E with β-micrating very low density lipoproteins (β-VLDL) and inhibits binding of β-VLDL to low density lipoprotein receptor-related protein. *J. Biol. Chem* **265:** 22453–22459.

CHAPTER 4

STRUCTURE AND EXPRESSION OF THE MOUSE APOLIPOPROTEIN C2 GENE.

M.J.V. Hoffer, M.M. van Eck, L.M. Havekes, M.H. Hofker, and R.R. Frants.

Genomics 1993, 17: 45-51

Structure and Expression of the Mouse Apolipoprotein C2 Gene

MARIËTTE J. V. HOFFER, MARGA M. VAN ECK, LOUIS M. HAVEKES,* MARTEN H. HOFKER,[1] AND RUNE R. FRANTS

*MGC-Department of Human Genetics, Leiden University, P O. Box 9503, 2300 RA Leiden, The Netherlands; and *IVVO-TNO, Gaubius Laboratory, P.O. Box 430, 2300 AK Leiden, The Netherlands*

Received December 30, 1992

Three cDNA clones containing the mouse apolipoprotein C2 (*Apoc2*) gene were isolated from a mouse liver cDNA library. The inserts from two cDNA clones were 500 bp in size while the insert from the third clone was unexpectedly large, 962 bp. All three clones contained a single open reading frame encoding apoC2. The exon–intron structure of the mouse *Apoc2* gene was determined by sequence analysis. Northern blotting and primer extension analysis of mouse RNA showed that the major liver transcript is 500 bp in size and is encoded by four exons. Transcripts for *Apoc2* were found in fetal liver, adult liver, intestine, and peritoneal macrophages. The largest cDNA clone, mAPOC2c4, contained an additional 440 bp at the 5′ end that are evolutionary conserved between man and mouse. These additional sequences are encoded by two exons located 5′ to the major liver start site. Although the larger transcript could not be detected by Northern blot analysis, products resulting from an upstream transcription initiation site were detected in the liver using RT-PCR analysis. The sizes of the RT-PCR products are consistent with alternative splicing. c 1993 Academic Press, Inc.

INTRODUCTION

Apolipoprotein C2 (apoC2) is a plasma lipoprotein associated with chylomicrons, very low density lipoproteins (VLDL) and high density lipoproteins (HDL). ApoC2 is required for the enzymatic activity of lipoprotein lipase (LPL), the enzyme responsible for hydrolysis of lipoprotein triglycerides to free fatty acids and monoglycerides (LaRosa *et al.*, 1970; Bier and Havel, 1970). Patients with apoC2 deficiency develop severe hypertriglyceridemia, which results in clinical phenotypes similar to LPL deficiency (Breckenridge *et al.*, 1978). Human apoC2, in its mature form, contains 79 amino acid residues with a molecular weight of 8.9 kDa (Jackson *et al.*, 1977) and is primarily synthesized in the liver (Marsh, 1976). A small fraction of human plasma apoC2 is proteolytically processed at the amino-terminal end (Fojo *et al.*, 1986). APOC2 nucleotide sequences have been determined from several species; human, monkey, dog, guinea pig, and rat (Mycklebost *et al.*, 1984; Whitted *et al.*, 1989; Datta *et al.*, 1987; Andersson *et al.*, 1991a,b). The human APOC2 gene consists of four exons separated by three introns and has been completely sequenced (Wei *et al.*, 1985; Das *et al.*, 1987). In man, the APOC2 gene is part of a gene cluster that is located on chromosome 19 and includes the APOE, APOC1, pseudo APOC1, and APOC2 genes (Smit *et al.*, 1988).

We have previously described the structure of the mouse *Apoe–c1–c2* gene cluster (Hoffer *et al.*, 1993). An unusually large *Apoc2* cDNA clone containing an extended 5′ end was isolated. We now describe the isolation of two additional mouse *Apoc2* cDNA clones and the characterization of the gene structure. The expression of the mouse *Apoc2* gene was investigated by Northern blotting and RT-PCR analysis, and the results suggest the presence of an additional upstream transcription start site.

MATERIAL AND METHODS

Isolation and sequencing of cDNA clones. A cDNA library from liver mRNA of an adult male BALB/c mouse (Clontech) was screened with a human APOC2 cDNA probe (Myklebost *et al.*, 1984). Inserts were isolated by PCR amplification using primers (a: 5′-GGTGGCGACGACTCCTGGAG-3′; b: 5′-GGTAATGGTAGCGACCGGCG-3′) flanking the *Eco*RI cloning site of λgt11. PCR analysis was performed as described for RT-PCR (see below) with the exception that 2.5 m*M* $MgCl_2$ was used in the 1× reaction buffer and the annealing temperature was 60°C. *Eco*RI-digested PCR products were subcloned into pUC19 or M13mp18/19 and sequenced using the T7 sequencing kit (Pharmacia). Individual clones were sequenced in both directions using the universal or reverse primer or specific synthetic oligonucleotides

Characterization of exon–intron structure Exons were located by hybridization with the mouse *Apoc2* cDNA probe (Hoffer *et al.*, 1993) For sequencing analysis, restriction enzyme fragments were subcloned into pUC19 and additional subclones were obtained using the Erase-a-base system (Promega) according the protocol of the manufacturer. All exons, exon–intron borders, and 5′ and 3′ flanking regions were sequenced using the T7 sequencing kit (Pharmacia) using universal or reverse primers and specific synthetic oligonucleotides.

mRNA isolation and characterization by Northern blotting analysis Peritoneal macrophages were induced by an intraperitoneal injection of 0.5 ml of sterile 0 8% glycogen. After 1 day, a booster was administered and 3 h later the cells were collected. Total cellular

[1] To whom correspondence should be addressed. Telephone: 31-71-276099. Fax: 31-71-276075.

0888-7543/93 $5.00

FIG. 1. Schematic representation of the cDNA sequence and sequencing strategy for the mouse *Apoc2* gene. (**A**) Schematic representation of the mouse *Apoc2* cDNA. The three cDNA clones are indicated below the *Apoc2* cDNA. The coding sequence of the cDNA is indicated by a shadowed box. (**B**) DNA sequencing strategy for the mouse *Apoc2* gene. Half-arrowheads indicate the direction of sequencing and vertical bars indicate that the universal or the reverse M13 sequencing primers are used. Open circles indicate the use of synthetic oligonucleotides as primer for sequencing analysis. The exons are indicated by solid boxes. Restriction sites include *Eco*RI (E), *Hin*dIII (H), *Kpn*I (K), and *Sac*I (S).

RNA from several tissues, including liver from a 12-day-old fetus, was isolated according to Chomczynski and Sacchi (1987). RNA samples were subjected to electrophoresis on a 1.2% (w/v) denaturing agarose gel containing 7.5% formaldehyde and transferred to a nylon membrane (Hybond N+, Amersham) according the protocol provided by the manufacturer. Hybridizations were performed at 55°C (Krumlauf, 1991). The cDNA probes were labeled using the random primer method (Multiprime labeling kit; Amersham) with [α-^{32}P]dCTP. The filters were washed twice in 2× SSC (SSC: 0.15 *M* NaCl, 15 m*M* sodium citrate), 0.5% SDS for 15 min at 65°C and twice with 1× SSC, 0.5% SDS for 15 min.

Primer extension analysis. Poly(A)$^+$ RNA was isolated using the polyATract mRNA isolation kit (Promega). A synthetic oligonucleotide (5′-GGACCTCATTTCCCAACATC-3′) was end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase and annealed to 2 μg poly(A)$^+$ RNA isolated from liver (Geliebter *et al.*, 1986). The primer was extended using AMV reverse transcriptase (Promega) as described by Geliebter *et al.* (1986). The extension product was run through a 6% sequencing gel. A sequencing reaction using the same oligonucleotide was carried out, and the sequence mAPOC2c4 served as a size standard to determine the size of the extension product.

RT-PCR analysis and direct sequencing. For synthesis of *Apoc2* cDNA, poly(A)$^+$ RNA from liver was reverse-transcribed using AMV reverse transcriptase. The reaction mixture (20 μl) contained 1 μg poly(A)$^+$ RNA, 0.5 μg oligo(dT), 1× reverse transcription buffer (Promega), 1 m*M* dNTPs, 20 U RNasin, and 8 U of AMV reverse transcriptase (Promega). This mixture was incubated at 42°C for 60 min, heated to 65°C for 5 min, quick chilled on ice, and subjected to PCR. Specific oligonucleotides were used to amplify mouse *Apoc2* cDNA between exons 1A, 1B, 1C, and exon 4 (exon 1A: 5′-CAGAGATAGAGAGGACCAAG-3′; exon 1B: 5′-TCCAGGGCTTTATGCAGACC-3′; exon 1C: 5′-CCTGCCACTACATTCAGGTC-3′; exon 4: 5′-GCTTTTGCTGTACATGTCCC-3′). Two microliters of cDNA mixture was amplified in a 50-μl reaction mixture containing 100 ng of each primer, 0.2 m*M* dNTPs, 1× reaction buffer (1.0 m*M* $MgCl_2$, 50 m*M* KCl, 10 m*M* Tris (pH 8.4), 0.2 mg/ml BSA), and 1 unit *Taq* DNA polymerase (Cetus Corp.) and overlayered with 50 μl of mineral oil. PCR reaction conditions were as follows: 1 min at 95°C, 1.5 min at 55°C, and 2 min at 72°C, for 32 cycles in an automated PCR machine (Biomed). After amplification, the samples were subjected to electrophoresis in 2% agarose gels. PCR samples used for direct sequencing were run on a 1% low-melting agarose gel and the band of interest was excised and heated to 65°C. A total of 2.5 μl of the melted agarose was mixed with 2 μl annealing buffer (T7 sequencing kit, Pharmacia) and 60 ng primer in a total volume of 14 μl. The mixture was heated to 95°C for 5 min and quick chilled on ice. Labeling and termination reactions were performed using the T7 sequencing kit according to the manufacturer (Pharmacia).

RESULTS

Isolation and Sequence Analysis of Mouse Apoc2 cDNA Clones

A mouse liver cDNA library was screened with a human APOC2 cDNA probe and three clones were obtained. Two *Apoc2* clones, mAPOC2c1 and c8, are both approximately 500 bp in size and map 19 bp apart at the 5′ end (Fig. 1A). Sequence analysis revealed that these clones contain exons 1–4 of the *Apoc2* gene, closely resemble other APOC2 cDNA clones (Andersson *et al.*,

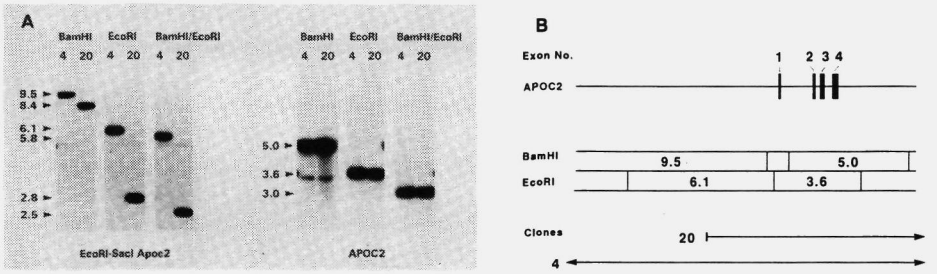

FIG. 2. Southern blot analysis of human APOC2 cosmids. (**A**) Human cosmid DNA (cosmid 4 and 20) was digested with *Bam*HI, *Eco*RI, or with both *Bam*HI and *Eco*RI as indicated above the autoradiogram. Blots were hybridized with a human APOC2 cDNA probe (**right**) or a 5′ *Eco*RI–*Sac*I fragment of the cDNA clone mAPOC2c4 (**left**). The sizes of the fragments are indicated in kb. In the left panel, the band in the lane of *Bam*HI-digested DNA of cosmid 20 includes 5 kb of the vector. The autoradiograms were exposed for 1 h at −80°C. (**B**) A restriction map of the human APOC2 gene for the restriction enzymes *Bam*HI and *Eco*RI. Cosmids used for Southern blotting are indicated beneath the map.

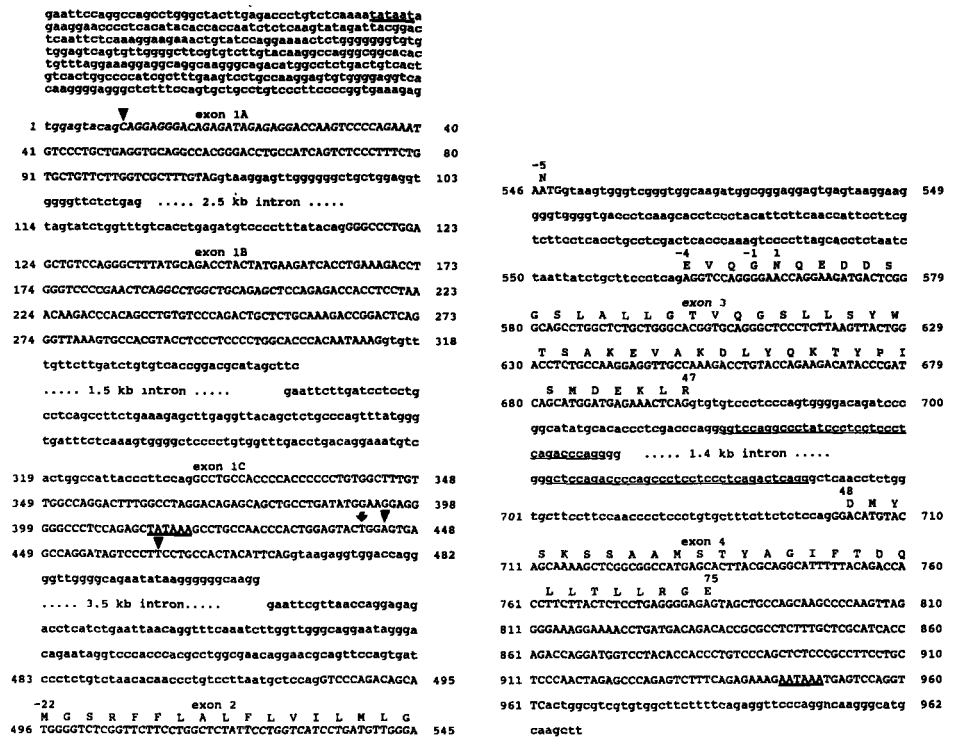

FIG. 3. Nucleotide sequence of the *Apoc2* gene. Exon nucleotide sequences and the flanking intron sequences are shown in uppercase and lowercase, respectively. The predicted amino acid sequence is indicated above the nucleotide sequence. The TATA box and polyadenylation signal are double underlined. The major liver transcription start site is indicated by an arrow above the sequence. Triangles above the sequence indicate the first nucleotide of the cDNA clones mAPOC2c4, c8, and c1, respectively. The sequence of the repeat unit of the VNTR within intron 3 is underlined.

1991b; Myklebost *et al.*, 1984), and show within the coding region 71% homology to the human APOC2 gene. One clone, mAPOC2c4, had the unexpected size of 962 bp. We have shown previously, that the 5′ portion (Fig. 1A) of this clone maps 5′ of the *Apoc2* gene (Hoffer *et al.*, 1993). This clone was completely sequenced (EMBL Accession No. Z15090) and found to contain the *Apoc2* gene and an additional 440 bp at the 5′ end.

The 5′ end of clone mAPOC2c4 was investigated for evolutionary conservation between man and mouse by Southern blot analysis using two human APOC2 cosmids, obtained during a previous study (Smit *et al.*, 1988). These two cosmids were digested with *Bam*HI, *Eco*RI, and *Bam*HI/*Eco*RI and hybridized with a human APOC2 probe or with a probe derived from a 200-bp *Eco*RI–*Sac*I fragment of the 5′ end of mAPOC2c4 (Fig. 2A). Hybridization with the human APOC2 cDNA probe revealed fragments containing exons 1–4 of the human gene. Using the 5′ end of mAPOC2c4, fragments located 5′ of the human APOC2 gene hybridized with this probe. Figure 2B shows a map of the two cosmids and the position of the human APOC2 gene.

Genomic Structure of the Mouse Apoc2 Gene

*Eco*RI fragments from cosmid mAPOE27 containing the *Apoe-c1-c2 gene cluster* (Hoffer *et al.*, 1993) were cloned and mapped for several restriction enzymes. A significant part of the genomic DNA of the mouse *Apoc2* gene and the additional genomic regions homologous to mAPOC2c4 were sequenced (EMBL Accession Nos. Z22213, Z22214, Z22215, Z22216, Z22217) according to the strategy outlined in Fig. 1B. The genomic sequence of *Apoc2* from exons 2–4 and a region around exon 1 are shown (Fig. 3). This gene region encodes both cDNA clones mAPOC2c1 and c8 and reveals a gene structure that is similar to the human APOC2 gene. A 195-bp 5′

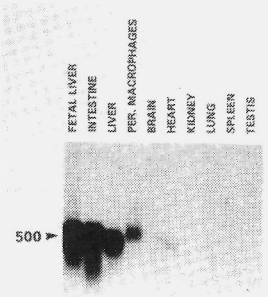

FIG. 4. Northern blot analysis. Total cellular RNA (10 μg) from several tissues, as indicated above each lane, was examined by Northern blot analysis. The Northern blot was probed with an *ApaI–Eco*RI fragment of mouse *Apoc2* cDNA containing exons 1C–4. The length of the mRNA in basepairs is indicated to the left. The autoradiogram was exposed overnight.

flanking sequence to mouse exon 1 shows 62% homology with the human promoter of APOC2. The 5′ portion of the larger transcript mAPOC2c4 included additional sequences located upstream of the major liver start site of *Apoc2*. The first 328 bp of the 5′ sequences of mAPOC2c4 are encoded by two separate exons (exon 1A and 1B) (Figs. 1B and 3). The upstream exon 1A is at least 123 bp long. Exon 1B is 205 bp long and contains the *Sac*I site. The remainder of the additional sequences are part of the liver promoter region, resulting in an enlarged exon 1 (exon 1C) at the 5′ end (Figs. 1B and 3). All exon–intron junctions, including exons 1A, 1B, and 1C, in the mouse *Apoc2* gene follow the GT/AG rule (Breathnach *et al.*, 1978).

Expression of the Apoc2 Gene

The size of the mouse *Apoc2* mRNA was investigated by Northern blotting of total liver RNA and RNA derived from several other tissues (Fig. 4). The length of the mRNA transcript in the liver was 500 bp. This size was confirmed by comparing mouse and human APOC2 transcripts from the liver (results not shown). *Apoc2* was also found to be expressed in intestine, peritoneal macrophages, and fetal liver. All tissues showed only transcripts of a similar size, i.e., 500 bp. A longer exposure of the Northern blots did not reveal a larger transcript, corresponding to mAPOC2c4.

The start site of the 500-bp liver transcript was analyzed by primer extension analysis. An oligonucleotide complementary to the 3′ region of exon 2 was annealed to mouse liver poly(A)$^+$ RNA and extended with AMV reverse transcriptase. The size of the extension product was 113 bp (Fig. 5) and predicts the location of the major start site of transcription in the liver to be within exon 1C (Fig. 3).

Characterization of Apoc2 mRNA Transcripts by RT-PCR

The transcript corresponding to the long cDNA, mAPOC2c4, could not be detected by Northern blotting. Therefore RT-PCR analysis was performed on liver poly(A)$^+$ RNA to demonstrate the existence of transcripts containing exons 1A, 1B, and 1C *in vivo*. After reverse-transcription with AMV reverse transcriptase to produce cDNA, PCR amplifications were performed using one of the three forward primers located in exon 1A, 1B, or 1C and the backward primer located in exon 4 (Fig. 6A). After 32 cycles the PCR product generated by the 1C/4 primer pair yielded a clear band on an ethidium bromide-stained agarose gel. With primer pairs 1A/4 and 1B/4 fragments of a lesser intensity were found. The identity of these bands was confirmed by Southern blot analysis with mAPOC2c4 (Fig. 6B) or specific oligonucleotides (not shown) as a probe. The size of the PCR product obtained with primer pair 1C/4 was 254 bp, as expected from the *Apoc2* gene structure. However, with primer pair 1A/4, 505- and 343-bp fragments were found instead of the expected 710-bp fragment. These two fragments were isolated, used as a template in a second PCR reaction with the same primers, and subsequently sequenced. The 505-bp fragment contained exons 1A, 1C, 2, 3, and 4 while the 343-bp fragment contained exons 1A, 2, 3, and 4 (Fig. 6A). The sequences showed that all exon junctions were correctly spliced (Fig. 6C). Primer pair 1B/4 also yielded two fragments (Fig. 6B). The size of the largest fragment corresponds to exons 1B, 1C, 2, 3, and 4 while the smaller fragment probably

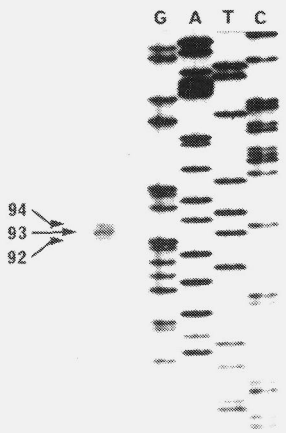

FIG. 5. Primer extension analysis of mouse *Apoc2* mRNA. Mouse poly(A)$^+$ liver RNA (1 μg) was annealed to a ^{32}P-labeled oligonucleotide specific for the 3′ end of exon 2 of *Apoc2*. The oligonucleotide was extended and analyzed as described under Material and Methods. A sequencing reaction was used in parallel lanes as a marker. The sequence was exposed for 1 h and the primer extension was exposed overnight.

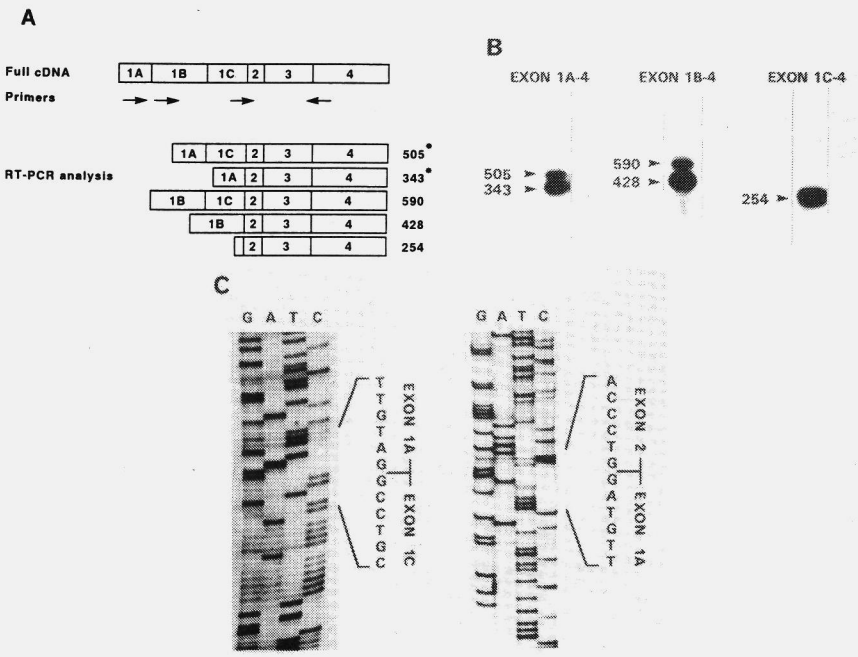

FIG. 6. RT-PCR analysis and sequencing of rare *Apoc2* transcripts. (**A**) A schematic representation of the *Apoc2* cDNA, the position of the primers used for the PCR analysis, and the alternative splicing products found by RT-PCR. Primers are indicated below the cDNA as arrows. The PCR products found with the RT-PCR are indicated below the cDNA. The size of the fragments is given to the right of each product. The fragment sizes indicated by an asterisk are confirmed by sequencing. (**B**) RT-PCR analysis with primer pairs as indicated above each lane. After PCR analysis and electrophoresis on an agarose gel, Southern blotting was performed with an *Apoc2* cDNA probe. The length of the fragments is indicated to the left of each lane. (**C**) Sequence of the fragments found with primer pair 1A/4. The alternative splicing boundaries are indicated to the right of each sequence. The left panel shows the 505-bp fragment and the right panel shows the 343-bp fragment.

lacks exon 1C (Fig. 6A). RNA from the other tissues expressing the 500-bp transcript did not show products with the 1A/4 and 1B/4 primer pairs, except for fetal liver, which gave results identical to those of liver.

DISCUSSION

Apoc2 Gene Structure

In the mouse, the *Apoc2* gene is contained within the *Apoe-c1-c2* gene cluster spanning 30 kb (Hoffer *et al.*, 1993). In the present study, we have isolated and characterized cDNA and genomic clones of the mouse *Apoc2* gene. The mouse cDNA encodes a protein of 97 amino acids residues including a signal peptide of 22 amino acid residues. In contrast to the human apoC2 protein (Jackson *et al.*, 1977), the deduced amino acid sequence of mouse *apoC2* is, like in the rat (Andersson *et al.*, 1991b), one residue shorter at the carboxyl terminus and lacks the conserved site for proteolytic cleavage of the proprotein. Interestingly, one of our cDNA clones, mAPOC2c4, contains a 962-bp insert due to an extended 5′ region compared to the other two cDNAs. However, the mAPOC2c4 clone lacks additional open reading frames. This result indicates that this long transcript still encodes the same protein. The cDNA clones mAPOC2c1 and c8 are transcribed from four exons. Exon 2 includes a part of the 5′ untranslated region and the signal peptide. Exons 3 and 4 include 79 amino acids of the protein coding sequences, and exon 4 includes the 3′ untranslated region, which is similar to the structure described for the human gene (Das *et al.*, 1987; Wei *et al.*, 1985). Primer extension analysis showed that the major liver start site of the 500-bp transcript is located in exon 1C. This start site coincides with a region of homology to the human APOC2 gene of 62% over 195 bp (Das *et al.*, 1987; Wei *et al.*, 1985). A classical TATA box resides 22 bp upstream of this site. In addition, the 5′ ends of the mAPOC2c1 and c8 clones are located just 3′ of the liver transcription start site.

Expression of mouse *Apoc2* was found in the liver and intestine as demonstrated in other species (Myklebost *et al.*, 1984; Andersson *et al.*, 1991b). In addition, we show that *Apoc2* is also expressed in peritoneal macrophages

like the other two genes, the mouse *Apoe* and the human APOC1, of this gene cluster (Basu *et al.*, 1981; Lauer *et al.*, 1988). High level expression of *Apoc2* was also detected in fetal liver mRNA.

Additional Sequences at the 5′ End

Sequence analysis of mAPOC2c4 and genomic clones shows the presence of two additional exons 5′ to the major liver start site. The exon–intron splice junctions of both exons follow the GT/AG rule (Breathnach *et al.*, 1978). Previously, a similar 5′ extended region was found in the cynomolgus monkey (Whitted *et al.*, 1989). Comparison of these 5′ additional sequences between monkey and mouse showed only a 71% homology for exon 1A. Southern blot analysis confirmed evolutionary conservation of the region 5′ of the major liver start site between man and mouse, indicating that these sequences may be of functional importance. However, the expression level of the large transcript in the tissues used for Northern blotting was below the detection level in all tissues investigated. Using the more sensitive RT-PCR analysis, transcripts were detected originating from a region 5′ of the major liver start site. Although the full-length transcript corresponding to mAPOC2c4 could not be demonstrated, other transcripts were found containing exon 1A but lacking exon 1B or exons 1B and 1C, indicating alternative splicing.

One explanation for the appearance of these larger transcripts could be the presence of upstream promoter sequences. The 5′ start site could be determined by a TATA box sequence located 315 bp 5′ of exon 1A. A similar situation was found for another protein produced mainly in the liver, the human α1-antitrypsin gene (Perlino *et al.*, 1987). Two different promoters, located 2 kb from each other, are active in macrophages or hepatocytes, resulting in two different transcripts. Since the liver contains Kupffer cells, endothelial cells, and fat cells, in addition to hepatocytes, we cannot exclude the fact that the 5′ transcripts originate from one of the latter less-abundant cell types. Also, the longer transcript may be present at a higher level during fetal gene expression, however, at an earlier stage than the fetal livers we have examined.

Another explanation for the low level transcripts can be offered by a recent finding of the rat ECL gene. The ECL gene may represent an additional gene within the APOE–C1–C2 gene cluster of the rat (Shen and Howlett, 1992). Comparison of the sequences of exons 1A and 1B of the *Apoc2* gene with the sequence of the rat ECL gene showed homologies of 87 and 95%, respectively. The orientation of the open reading frame of ECL implies that the transcription would be in the opposite direction with respect to *Apoc2*. Transcription of this gene may cause an "open chromatin" structure, clearing access of transcription factors to a cryptic *Apoc2* promoter region.

The mouse *Apoc2* gene has been characterized at the cDNA and genomic levels. Here we show that the major 500-bp *Apoc2* transcript in the liver is encoded by four exons. A rare transcript with additional sequences at the 5′ end has also been detected that is conserved among several species. The isolation of a larger cDNA clone and the detection of alternative splicing products of a larger transcript by RT-PCR indicate that transcription originates from a region 5′ of the major liver start site in the mouse *Apoc2* gene. Since all PCR products and the cDNA clone containing additional 5′ sequences were spliced properly, it is unlikely that they result from artifacts. Further studies are needed to unravel the origin and possible biological function of the apparently untranslated 5′ end of these larger *Apoc2* transcripts.

ACKNOWLEDGMENTS

The authors acknowledge Dr. K. Willems van Dijk for his helpful discussions and Dr. M Kielman for providing fetal liver mRNA. This work was supported by HGO-TNO (900-539-052) and the Royal Netherlands Academy of Arts and Science (M.H.H.).

REFERENCES

Andersson, Y., Thelander, L., and Bengtsson-Olivecrona, G. (1991a). Demonstration of apolipoprotein CII in guinea pigs *J. Biol Chem* **266:** 4074–4080.

Andersson, Y., Thelander, L., and Bengtsson-Olivecrona, G. (1991b). Rat apolipoprotein C-II lacks the conserved site for proteolytic cleavage of the pro-form. *J Lipid Res* **32:** 1805–1809.

Basu, S. K., Brown, M. S., Ho, Y. K., Havel, R. J., and Goldstein, J. L. (1981). Mouse macrophages synthesize and secrete a protein resembling apolipoprotein E *Proc Natl. Acad. Sci USA* **78:** 7545–7549.

Bier, D. M., and Havel, R. J. (1970). Activation of lipoprotein lipase by lipoprotein fractions of human serum. *J. Lipid Res* **11:** 565–570.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon–intron boundaries. *Proc. Natl Acad Sci. USA* **75:** 4853–4857

Breckenridge, W. C., Little, J. A., Steiner, G., Chow, A., and Poapst, M. (1978) Hypertriglyceridemia associated with deficiency of apolipoprotein C-II. *N Engl J Med* **298:** 1265–1273.

Chomczynski, P., and Sacchi, N. (1987). Single step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction. *Anal Biochem.* **162:** 156–159.

Das, H. K., Jackson, C. L., Miller, D. A., Leff, T., and Breslow, J. L. (1987) The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron. *J Biol. Chem* **262:** 4787–4793.

Datta, S., Li, W-H., Ghosh, I., Luo, C-C., and Chan, L. (1987). Structure and expression of dog apolipoprotein C-II and C-III mRNAs. *J. Biol. Chem* **262:** 10588–10593.

Fojo, S. S., Taam, L., Fairwell, T., Ronan, R., Bishop, C., Meng, M. S., Hoeg, J. M., Sprecher, D. L., and Brewer, H. B., Jr. (1986). Human preproapolipoprotein C-II *J. Biol. Chem.* **261:** 9591–9594.

Geliebter, J., Zeff, R. A., Melvold, R. W., and Nathenson, S. G. (1986). Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequencing analysis: K^{bm9} and K^{bm6}. *Proc Natl. Acad Sci. USA* **83:** 3371–3375.

Hoffer, M. J. V., Hofker, M. H., van Eck, M. M., Havekes, L. M., and Frants, R. R. (1993) Evolutionary conservation of the mouse apolipoprotein e–c1–c2 gene cluster: Structure and genetic variability in inbred mice. *Genomics* **15:** 62–67.

Jackson, R. J., Baker, H. N., Gilliam, E. B., and Gotto, A. M., Jr. (1977). Primary structure of very low density apolipoprotein C-II of human plasma *Proc. Natl Acad Sci. USA* **74:** 1942–1944.

Krumlauf, R. (1991). Northern blot analysis of gene expression. *In*

"Gene Transfer and Expression Protocols" (E. J. Murray, Ed.), pp. 307–360, Humana Press, Clifton, NJ.

LaRosa, J. C., Levy, R. I., Herbert, P., Lux, S. E., and Frederickson, D. S. (1970). A specific apoprotein activator for lipoprotein lipase. *Biochem. Biophys. Res. Commun.* **41:** 57–62.

Lauer, S. J., Walker, D., Elshourbagy, N. A., Reardon, C. A., Levy-Wilson, B., and Taylor, J. M. (1988). Two copies of the human apolipoprotein C-I gene are linked closely to the apolipoprotein E gene. *J. Biol. Chem.* **263:** 7277–7286.

Marsh, J. B. (1976). Apoproteins of the lipoproteins in a nonrecirculating perfusate of rat liver. *J. Lipid Res* **17:** 85–90.

Myklebost, O., Williamson, B., Markham, A. F., Myklebost, S. R., Rogers, J., Woods, D. E., and Humphries, S. E. (1984). The isolation and characterization of cDNA clones for human apolipoprotein CII. *J. Biol Chem* **259:** 4401–4404.

Perlino, E., Cortese, R., and Ciliberto, G. (1987). The human α_1-antitrypsin gene is transcribed from two different promoters in macrophages and hepatocytes. *EMBO J* **6:** 2767–2771.

Shen, P., and Howlett, G. J. (1992). Two coding regions closely linked to the rat apolipoprotein E gene: Nucleotide sequences of rat apolipoprotein C-1 and ECL cDNA. *Arch. Biochem. Biophys.* **297:** 345–353.

Smit, M., van der Kooy-Meys, E., Frants, R. R., Havekes, L. M., and Klasen, E. C. (1988). Apolipoprotein gene cluster on chromosome 19. *Hum. Genet.* **78:** 90–93.

Wei, C-F., Tsao, Y-K., Robberson, D. L., Gotto, A. M., Jr., Brown, K., and Chan, L. (1985). The structure of the human apolipoprotein C-II gene. *J. Biol. Chem* **260:** 15211–15221.

Whitted, B. E., Castle, C. K., Polites, H. G., Melchior, G. W., and Marotti, K. R. (1989). Purification, cloning and nucleotide sequence determination of cynomolgus monkey apolipoprotein C-II: Comparison to the human sequence. *Mol. Cell. Biochem.* **90:** 69–79.

CHAPTER 5

THE APOLIPOPROTEIN C2 LINKED GENE (*Acl*) GENE: DETECTION OF A NOVEL GENE WITHIN THE APOLIPOPROTEIN E-C1-C2 GENE CLUSTER IN THE MOUSE.

M.J.V. Hoffer, M.M. van Eck, L.M. Havekes, M.H. Hofker, and R.R. Frants.

Submitted

THE APOLIPOPROTEIN C2 LINKED (*Acl*) GENE: A NOVEL GENE WITHIN THE MOUSE APOLIPOPROTEIN E-C1-C2 GENE CLUSTER

Mariëtte J.V. Hoffer, Marga M. van Eck, Louis M. Havekes#, Rune R. Frants and Marten H. Hofker

MGC-Department of Human Genetics, Leiden University, P.O.Box 9503, 2300 RA Leiden, The Netherlands.#TVVO-TNO, Gaubius Laboratory, P.O.Box 430, 2300 AK Leiden, The Netherlands.

SUMMARY

The apolipoprotein E, C1 and C2 genes are contained within a gene cluster in man. Previously, we have shown that this gene cluster has been evolutionarily conserved in mouse. During the characterization of the mouse *Apoc2* gene, evolutionarily conserved and transcribed sequences were found 5' of the *Apoc2* gene. In this study, we show that these 5' additional sequences represent a novel gene within the gene cluster, designated the apolipoprotein C2 linked gene (*Acl*). The *Acl* gene is located 2 kb 5' to the *Apoc2* gene and has the same transcriptional orientation as all the other genes within the *Apoe-c1-c2* gene cluster. We have sequenced the entire mouse *Acl* gene on cDNA and genomic level. The gene consists of 3 exons, spanning a region of approximately 3.6 kb. The *Acl* gene is highly expressed in the liver with a transcript of 473 bp in size and encodes a putative protein of 124 amino acid residues.

INTRODUCTION

Apolipoproteins are structural constituents of lipoprotein particles and exhibit specialized functions in lipoprotein metabolism, serving as enzyme cofactors and as ligands for receptor-mediated uptake of lipoprotein particles (Breslow, 1988). Seven of these well-characterized apolipoproteins (APO), APOA1, A2, A4, C1, C2, C3, and E, are members of a multigene gene family because of structural similarities (for a review, see Li et al., 1988). All these genes consist of 4 exons with the exception of APOA4 which consists of only 3 exons. In addition, the coding regions are composed

of tandem repeats of 11 codons, indicating that these genes have evolved through duplications of a primordial gene (Luo et al., 1986; Li et al., 1988).

The apolipoprotein genes have become dispersed in mammalian genomes. In man six of these genes are found in two gene clusters, the APOA1-C3-A4 gene cluster on chromosome 11 (Bruns et al., 1984; Karathanasis, 1985) and the APOE-C1-C2 gene cluster on chromosome 19 (Lusis et al., 1986). The APOA2 gene is found as single gene on chromosome 1 (Lackner et al., 1984). These loci are contained in syntenic linkage groups in the mouse, which implies that all these genes have followed the same evolutionary pathway in humans and rodents (Lusis and Sparkes, 1989).

Recently, we have characterized the *Apoe-c1-c2* gene cluster in the mouse (Hoffer et al., 1993a) and demonstrated a similar gene organization as in man with the exception that the mouse gene cluster does not contain a pseudo *Apoc1* gene. Characterization of the genomic structure of the individual mouse genes showed that the mouse *Apoe* (Horiuchi et al., 1989) and *Apoc1* (Hoffer et al., 1993c) genes are similar to the corresponding human genes. Also the structure of the mouse *Apoc2* gene is comparable with its human counterpart. Interestingly, in addition to the normal sized cDNA clones we also found a relatively large *Apoc2* cDNA clone containing additional 5' sequences (Hoffer et al., 1993b). However, the corresponding transcript could not be detected by Northern blot analysis suggesting that it represented a low abundant mRNA species. In the present study we show that these additional 5' sequences of the large *Apoc2* cDNA clone are part of a novel gene within the *Apoe-c1-c2* gene cluster. This novel gene, designated the apolipoprotein C2 linked (*Acl*) gene, has been characterized further and the tissue expression has been determined.

MATERIALS AND METHODS

Isolation and sequencing of cDNA and genomic clones.

An *Eco*RI-*Sac*I restriction fragment of the 5'-end of the long *Apoc2* cDNA clone mAPOC2c4 was subcloned and used as a probe to screen a liver cDNA library (BALB/c) in λgt11 (Clontech). Inserts of positive clones were obtained by PCR using primers derived from the the nucleotide sequence around the cloning *Eco*RI site of lambda GT11 (Hoffer et al., 1993b) and either sequenced directly (Circumvent Thermal Cycle dideoxy DNA sequencing kit, Biolabs) or subcloned into the pT7Blue (R) T-vector (Novagen), followed by sequencing using the T7 sequencing kit (Pharmacia). Individual clones were sequenced in both directions using universal primers or specific oligonucleotides. A 4.5 kb *Eco*RI fragment derived from the genomic cosmid clone, mAPOE27 (Hoffer *et al.*, 1993a) was subcloned into pUC19 for sequencing. Sequencing

reactions were performed using the T7 sequencing kit (Pharmacia) and specific synthetic oligonucleotide primers (Fig. 2).

Northern blot analysis.

Total cellular RNA from several mouse tissues, including liver from a 12-day-old fetus, was isolated according to Chomczynski and Sacchi (1987). Peritoneal macrophages were isolated after induction by intraperitoneal injection of 0.5 ml sterile 0.8% glycogen. After one day a booster was administered and 3 hours later the cells were collected. RNA samples were subjected to electrophoresis on a 1.2% (w/v) denaturing agarose gel containing 7.5% formaldehyde and transferred to a nylon membrane (Hybond N^+, Amersham) according to the protocol provided by the manufacturer. Hybridizations were performed at 55°C in a solution containing 50% formamide (Krumlauf, 1991). The cDNA probes were labelled using the random primer method (Multiprime labelling kit, Amersham) with [α-^{32}P]-dCTP. The filters were washed twice in 2 x SSC (SSC: 0.15 M NaCl, 15 mM sodium citrate), 0.1% SDS for 15 min at 65°C and twice with 1 x SSC, 0.1% SDS for 15 min.

Computer analysis.

Sequences from the Genbank or from the EMBL DNA data banks were compared using the "fasta" algorithm of the Genetics Computer Group Sequence Analysis Software Package Version 7.2 (Madison, Wisconsin, USA).

RESULTS

Isolation of cDNA clones.

During the characterization of the mouse *Apoc2* gene, an unusually large cDNA clone, mAPOC2c4, was isolated (Hoffer et al., 1993b). This clone contained additional 5' sequences compared with the most abundant, normal sized *Apoc2* liver transcript. These additional sequences are encoded by two exons located 5' of the major *Apoc2* liver start site and are evolutionary conserved (Hoffer et al., 1993b) indicating that these sequences could be of functional importance. Further studies were performed to unravel the origin of these additional 5' sequences.

Therefore, a liver cDNA library, primed with random primers and oligo(dT), was screened with a 0.2 kb *Eco*RI-*Sac*I fragment as probe, containing only a part of the additional 5' sequences of the large *Apoc2* cDNA clone mAPOC2c4. The screening revealed one clone, mACLc1, with an insert of 431 bp (Fig. 1). Sequence analysis revealed a polyA tail, indicating that this clone was complete at its 3' end. This clone contains also sequences not corresponding with the 5' additional sequences of the large *Apoc2* cDNA clone. At the 5' end the sequence was incomplete compared with the sequences of the *Eco*RI-*Sac*I fragment, used as a probe for the cDNA library

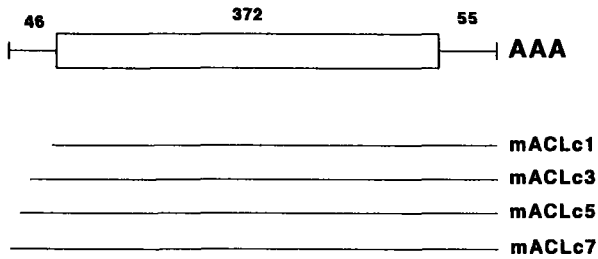

Fig. 1 Schematic representation of the cDNA sequence representing the *Acl* gene. The cDNA clones are indicated beneath the map. The encoding region is indicated by a box. The sizes (in nucleotides) of respectively the 5' untranslated region, the coding region, and the 3' untranslated region are indicated above the map.

screening. Hence, another aliquot of the cDNA library was screened with the insert of the cDNA clone mACLc1. The inserts of 7 additional clones were obtained by PCR and the sequence of their 5' ends was determined by direct sequencing. Three of these 7 clones (see Fig. 1) had an extended 5' end compared with mACLc1 and map 12 bp from each other. A potential full-length clone of 473 bp (mACLc7) was analysed in more detail.

The sequence of the mACLc7 cDNA clone has one major open reading frame. Postulating the first ATG codon as an initiating codon, the sequence consists of a 5'-untranslated region of 46 nucleotides, followed by 372 nucleotides encoding a putaive protein of 124 amino acids. The 3'-untranslated region of 55 nucleotides contains a polyadenylation signal located 15 nucleotides upstream from the polyadenylation site (Fig. 1). These results indicate that the cDNA clones represent a novel gene within the *Apoe-c1-c2* gene cluster and therefore, we have analyzed this apolipoprotein C2 linked (*Acl*) gene further.

The genomic structure of the Apolipoprotein C2 linked (Acl) gene.

A 4.5 kb genomic *Eco*RI fragment containing the *Acl* gene was cloned and mapped for several restriction enzymes. A part of the genomic sequence of *Acl* had been obtained previously during the characterization of *Apoc2* (Hoffer et al., 1993b). This sequence has been completed by sequencing the remaining gaps using synthetic primers according to the strategy outlined in Figure 2. The genomic sequence (Fig. 3) shows that the *Acl* gene consists of 3 exons and spans a region of 3.6 kb (Fig. 2 and 3). The position of the first nucleotide of the different cDNA clones relative to the genomic sequence is given for the 3 largest cDNA clones (Fig. 3). The first intron of 2570 bp

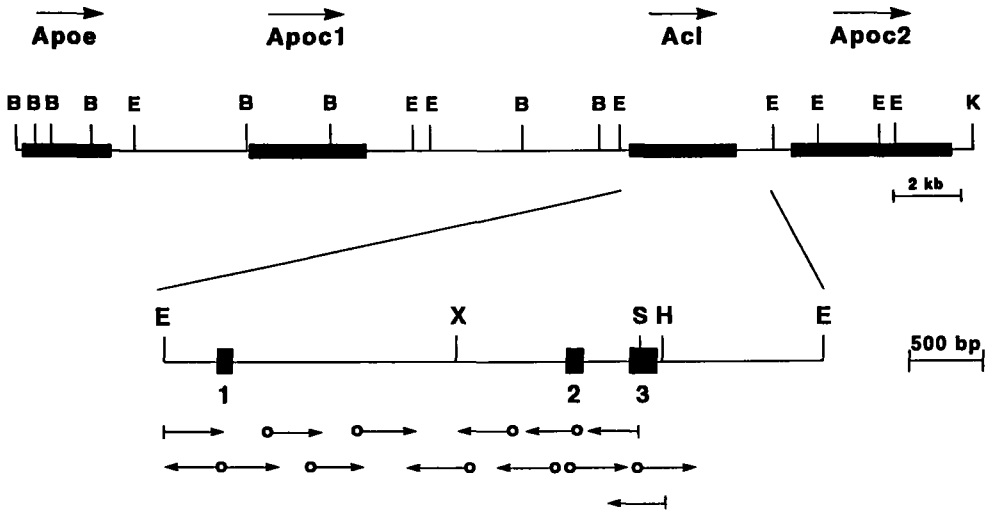

Fig. 2 Restriction site map, genomic organization and sequence strategy of the mouse *Acl* gene. At the top, a map of the mouse gene cluster containing the *Apoe*, *Apoc1*, *Acl*, and *Apoc2* genes, is shown. The genes are indicated by filled boxes and the transcriptional orientation is shown by arrows above the map. Below the overall map, an expanded map of the mouse *Acl* gene shows the exon-intron structure, indicating exons by filled boxes. Sequencing strategy is shown at the bottom of the figure. Arrows indicate the direction of sequencing. The nucleotide sequence was determined with specific oligonucleotides (open circles) or with universal primers (vertical bars). Restriction enzymes are indicated as follows: B, *Bam*HI; E, *Eco*RI; K, *Kpn*I; S, *Sac*I; X, *Xho*I.

interrupts the codon specifying amino acid 26 of the deduced amino acid sequence. Intron 1 contains 5 regions which are homologous to *Alu* like sequences. The second intron, interrupting the codon specifying amino acid 70, is 295 bp long and contains 3½ repeat units. These repeat units show a high homology with the VNTR found in the *Apoc2* gene but are in the opposite orientation (Hoffer et al., 1993a). All exon-intron junctions in the mouse *Acl* gene follow the GT/AG rule (Breathnach et al., 1978).

Northern blot analysis.

The *Acl* gene shows expression in the liver (Fig. 4) . Other tissues do not express Acl including those showing abundant expression of *Apoc2* such as intestine (Fig. 4) and peritoneal macrophages (not shown). The estimated size is in accordance with the other data found with the cDNA clones.

```
gaattccaggccagcctgggctacttgagaccctgtctcaaatataatagaaggaacccctcacataca   70
ccaccaatctctcaagtatagattacggactcaattctcaaaggaagaaactgtatccaggaaaactctg  140
gggggtgtgtggagtcagtgttggggcttcgtgtcttgtacaaggccagggcggcacactgtttaggaa  210
aggaggcaggcaagggcagacatggcctctgactgtcactgtcactggccccatcgctttgaagtcctgc  280
caaggagtgtggggaggtcacaaggggagggctctttccagtgctgcctgtcccttccccgtgaaagagt  350
                                                       1
 *       * *                                           M  S  L  L  R  C  R  P
gGAGTACAGCAGGAGGGACAGAGATAGAGAGGACCAAGTCCCCAGAAATGTCCCTGCTGAGGTGCAGGCC  420
                                                        25
 R  D  L  P  S  V  S  L  S  V  L  F  L  V  S  F  V
ACGGGACCTGCCATCAGTCTCCCTTTCTGTGCTGTTCTTGGTCAGCTTTGTAGgtaaggagttgggggc  490
tgctggaggtggggttctctgatggggcacgactgtggcttcatcagtgggtcagtagccatgtaagagg  560
aacgtgacacacacatgcgtacacacacatgcacacacagtcatgcctgcctctgcctcccaagtgctgg  630
                              Alu
gcttaaaggcgtgcgccaccaccgcccggctgtgtgggtttttttagagtgtacatgagtacactgtag  700
ctatcttcacacacaccagaaggatgccattacagatagatggttgtgagcctacatgtggttgctaaga  770
attgaactcattgtatctcaatgctcttaactctgagccatctctctagccccaccccgacactcagttt  840
ttaaaagtacagaaaactgggacccatgagatgactcagtgggtcaaggtgcatgccacaaaagcctggg  910
accggaggttagtccctggaactcatgtggtggaagggaaaacaggcttcaaaaggttgtcctctgacct  980
ccacaggaggatcatgagacacagaggacctcacacacagaaacacatattacacatacacatatacaca 1050
cataaatgagttaaaacttttattcttatttatttattctactttattttatgtgtgttggtgtctt 1120
                              Alu
gcctacacgttagactgtttgagtgtgtcagacccccctagaactaggttacagacagttgtgagctgcca 1190
tgtgggtgctgagaactgaagaactgggtcctctgctgggagagcactgagtcctcttaatgagcagcca 1260
gtgcccttaacagctgagctatctctctagccctttttcttttaattttcagacagaatggcctggagt 1330
tatagcttagcatgcatgaggcctttggttcaacatccagtaccattaaaataaataaattaattaaaat 1400
                              Alu
acaattgccgggcagtggtggtgcacgcctttaatccccagcacttgggaggcagaggtaggtggatttc 1470
tgagttcggggccagcctggtctacagagtgagttccaggacagccagtactacaccgagaaactctgtc 1540
tcggaacccccccccccccaaaaaattccccatgtatctcccaagacagtgcttaggggatgcccagccg 1610
gcaactcccgtgtactcctcattggcttagctcagtgcccctttcacagaacaggaaatagtgaaagtca 1680
ctgatgggacttcagaagggtcttggaactagagttcatgtgtggtctgggtgtgtttcctggaacccca 1750
                              Alu
gtgcttgggaggtggaggcaggaggatcaggagttcaaggctaatctcagctatgtaggaagtttctaga 1820
caagcagagaatgtggctcggtggagcttaggtggcatgcccaaacccgggttctagccctatcctgtca 1890
taaaacagacatggggttagacacccatcagcctgacatttggaggtggagtctagaggatcagcagtt 1960
```

```
taatggcagaggtcatccttgacactcgaggccatctggaactataagatggccatgtcatcaaaataac 2030

aacaacaacaataataacaataacaataggctagagggatgtctcagtggtagccccatagttaagggca 2100

tgcactgctcttacagaggaccagagttcagttcccagcctcacagctgcctgtaactccagctgcaggg 2170

actctgacacctttgctcttcacaggcacaggcagtacacttgtttatgtgcccgcagacaggaccatac 2240
                                  Alu
acacaggataaaaaataaaatagtgctggagagatggctcagtggttaagagcactggctgctcttccag 2310

aggtcctgagttcaattcccagcaaccacatggtgcctcacaaccatctgtggtgggcctgtctaatgga 2380

tctatgccctcttctggtgtgtctgaagatagctacagtgtactcacatacatgaaataagtaaataagt 2450

aaataaattcaaaaaaatgtttaaaaaaaaaaaacaaacaacacaagcatgctcatgtatacttgtgtgc 2520

caggtagtatctgggcgtgagtctctacctagttgtgtgggtgctggagcaggaggaatgtgtgtgtgtg 2590

tgtgtgtgtgtgtgtgtgtgtgtgtcccagagggcagagggttaggtggttagaggagagaggc 2660

accaggtgggaagtggagttgcacactgagcaagaaaatcccagtagacagagaaacaagggagggacca 2730

gcagtgaggaagagacagatcaaatggtggagagacggttgtggtgtcagaggcctaataggccaggatg 2800

tgcaaagtagtgcaagagactgggtactttcttcttaactatgcccctgtgttgtataataccttcaaa 2870

aaggacatcatggtgccctcttctacaggatgggaaactgaatctaggccaggtgaaatcaggtgtggcc 2940

tggaataaggggtgatgtgggtgtctaggtacaaggatgagcccacctgtccccaagagaggaatggggt 3010
                                  26
                                  A S M S T E S L S P T P
cctgagatctcaagcccctctgtggtttccccagCATCCATGTCTACAGAAAGCCTGAGCCCCACGCCTG 3080
G P E S S R W S L V R A R V L E M V E P L V T R
GCCCTGAGAGCAGCCGCTGGAGCCTGGTAAGGGCCAGGGTGTTGGAGATGGTGGAGCCACTGGTGACCAG 3150
                  70
  T R D R W Q W F W
AACCAGGGACAGATGGCAGTGGTTCTGgtgagacaatcctagtgcttggctgggagggaacaccagagtt 3220

agaggggaaagagagcctgtggctagatcctgggtctgagggaggagggctggggcctggagccctgag 3290

tctgaaggaggatggctggggcctgaactcttgggtctgagggaggagagctggggcctggacccctggg 3360

cctaaaggaggagtacaaagactctgaatcttgagtctaagaagaaggtcataagtcgggattgaacccc 3430
                                  71
                                  G P G A V Q G F M
tgtagtatctggtttgtcacctgagatgtccccttatacagGGGCCCTGGAGCTGTCCAGGGCTTTATG 3500
 Q T Y Y E D H L K D L G P R T Q A W L Q S S R
CAGACCTACTATGAAGATCACCTGAAAGACCTGGGTCCCCGAACTCAGGCCTGGCTGCAGAGCTCCAGAG 3570
                                                  124
D H L L N K T H S L C P R L L C K D R T Q G
ACCACCTCCTAAACAAGACCCACAGCCTGTGTCCCAGACTGCTCTGCAAAGACCGGACTCAGGGTTAAAG 3640

TGCCACGTACCTCCCTCCCCTGGCACCCACAATAAAGGTGTTTGTTCTTGatctgtgtcaccggacgcat 3710

agcttctcttgaggccagaggtcaaccttccgtcatatctaggagttaataagtggaaaagttgtggcct 3780

ggctggggaggacacaaggaaagctttaaaggaggggaccaaagtgataaacctccccctgccccacagg 3850

ttcctggcttgatgagaacaggagacttcca 3920
```

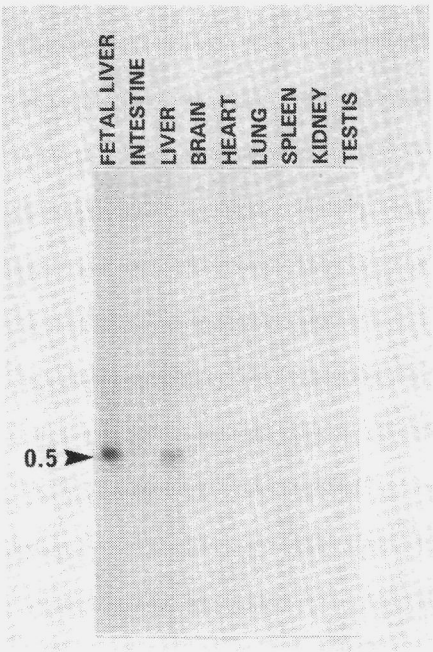

Fig. 3 (previous page) Nucleotide and deduced amino acid sequence of the mouse *Acl* gene. Exon nucleotide sequences are shown in upper case. The deduced amino acid sequence, in one letter code, is indicated above the nucleotide sequence. Asteriks above the sequence indicate the first nucleotide of the cDNA clones mACLc3, mACLc5, and mACLc7, respectively. The polyadenylation signal is doubly underlined. The VNTR repeat units are indicated by an arrow above the sequence and the *Alu* like sequences are labeled. (EMBL Accession No. Z24722).

Fig. 4 Northern blot analysis. Total cellular RNA (10μg) isolated from several tissues as indicated above each lane, was examined by Northern blot analysis. The Northern blot was probed with a fragment of the mouse *Acl* cDNA containing exons 1-3. The length of the mRNA in kilobase is indicated at the left.

DISCUSSION

Recently, we have isolated the mouse *Apoe-c1-c2* gene cluster and characterized the *Apoc1* and *Apoc2* genes in more detail (Hoffer et al., 1993a, b, and c). During the screening of a cDNA library an unusual large *Apoc2* cDNA clone, mAPOC2c4, was isolated, which exceeds the size of the regular liver transcript by 440 bp in size at the 5'end (Fig. 5). The screening of a cDNA library with these 5' sequences revealed cDNA clones encoding a novel gene located 5' of the *Apoc2* gene which was designated the apolipoprotein C2 linked (*Acl*) gene (Fig. 5). The finding of this novel gene 5' of *Apoc2* implies, that the intergeneous transcripts containing both genes could have been arisen from "read through" events beyond the polyadenylation signal of the *Acl* gene. Similar types of transcripts, generated by alternative splicing involving the APC and SRP19 genes have been described earlier (Horii et al., 1993). However, the very low abundance of the intergeneous transcripts and the lack of an additional open reading frame suggest that these transcripts containing the *Acl* and *Apoc2* genes are apparently artificial rather than of functional importance.

Using Southern blot analysis, we have shown in earlier studies that this *Acl* gene is present in the human APOE-C1-C2 gene cluster as well (Hoffer et al., 1993b). In

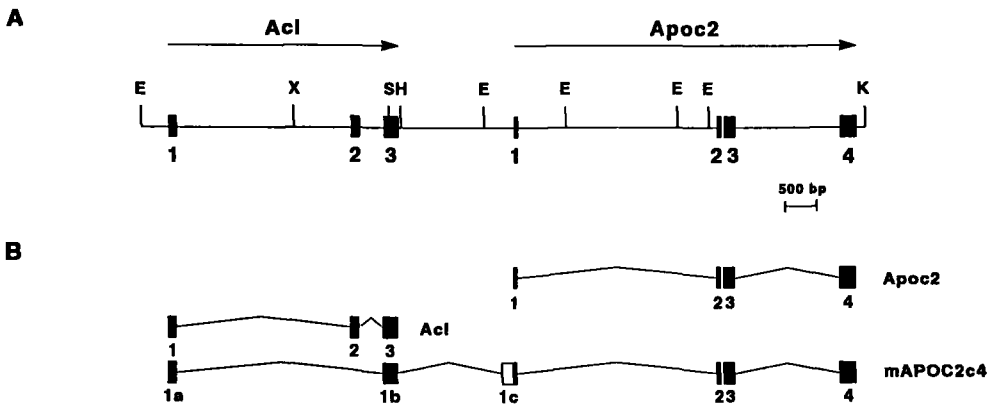

Fig. 5 Structure of the *Acl* and *Apoc2* genes. Below the map, a schematic representation of the transcripts for each individual gene and the intergeneous transcript, mAPOC2c4. Exons 1A and 1B of mAPOC2c4 are identical with exons 1 and 3 of the *Acl* transcript. Exons are indicated by filled boxes. The open box indicates additional 5' sequences in exon 1c of mAPOC2c4 compared to exon 1 of *Apoc2*.

addition, two previously reported cDNA sequences are homologous to the *Acl* gene. The first 250 bp of a APOC2 cDNA clone from cynomolgus monkey show 68% homology with the mouse *Acl* gene and resembles the intergeneous transcript represented by the mAPOC2c4 cDNA clone of mouse *Apoc2* (Whitted et al., 1989; Hoffer et al., 1993b). In addition, a homologous cDNA (ECL gene) of the rat has been described (Shen and Howlett, 1990). These investigators proposed a different transcriptional orientation leading to a predicted amino acid sequence which is very different from our data. Applying the other translational orientation the deduced amino acid sequence although interrupted by a stop codon, shows a high homology with Acl (Fig. 6). Comparison of the deduced amino acid sequence encoded by the *Acl* cDNA with the deduced amino acid sequences of the first part of monkey APOC2 and with that found for the ECL gene, when translated in the opposite orientation, shows discrete regions with a high degree of conservation (Fig. 6).

Acl was found to be expressed primarily in liver. Other tissues, including those showing abundant expression of *Apoc2*, such as intestine and peritoneal macrophages do not express *Acl*. The length of the mouse *Acl* transcript was approximately 480 bp and almost similar in size with the *Apoc2* transcript (Hoffer et al., 1993b). This explains that Northern blot analysis using a cDNA clone, containing *Acl* as well as

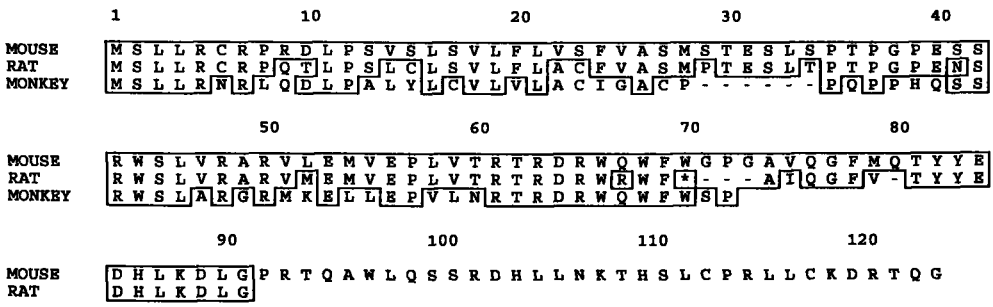

Fig. 6 Comparison of the deduced amino acid sequence of the mouse *Acl* and partial sequences of rat and monkey homologues. Breaks indicated by (-) are introduced to optimize the homology among these three sequences. Asterik indicates a stop codon. The numbering of the sequences is based on the start of mouse *Acl*. Sequence data for rat are from Shen and Howlett (1992), and for monkey from Whitted *et al.* (1989).

Apoc2 as a probe did not reveal a larger or additional transcript.

The presence of the *Acl* gene within the *Apoe-c1-c2* gene cluster makes this gene a candidate to be a potential new member of the apolipoprotein multigene family. It has been suggested that several apolipoproteins are derived form a common evolutionary precursor (Luo et al., 1986). In this model the apolipoproteins consist of three 11-amino acid long repeats and of a variable number of 22-amino-acid long sequences that are found at the COOH-terminal part of the protein. We have examined the *Acl* sequence for the presence of internal duplications at the DNA level. Computer-aided analysis showed that the DNA sequence contains one repeated DNA region that has a high degree of homology (results not shown).

In addition, we have analysed the *Acl* gene on protein level for the presence of structural features indicative for apolipoproteins. Two regions with amphiphatic helices were found which are homologous with amphipatic helices from known apolipoproteins (Hans De Loof, personal communications). The helical hydrophobic moment shows a minimium at the boundary of exons two and three which is also found for the apolipoproteins of the multigene family. However, further analysis on the function of this protein are needed to unravel whether this gene encodes a functional apolipoprotein. Currently, polyclonal antibodies are produced to perform further studies at protein level.

ACKNOWLEDGEMENTS

The authors acknowledge Dr. H. De Loof, Dr. P. de Knijff and Dr. R. Fodde for their helpful discussions. This work was supported by HGO-TNO (900-539-052) and the Royal Netherlands Academy of Arts and Science (M.H.H.).

REFERENCES

Breathnach R, Benoist C, O'Hare K, Gannon F, and Chambon P (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc Natl Acad Sci USA 75: 4853-4857.

Breslow JL (1988). Apolipoprotein genetic variation and human disease. Physiol Rev 68:85-132.

Bruns GAP, Karathanasis SK, and Breslow JL (1984). Human apolipoprotein A-I-C-III complex is located on chromosome 11. Arteriosclerosis 4: 97-102.

Chomczynski P, and Sacchi N (1987). Single step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction. Anal Biochem 162: 156-159.

Das HK, Jackson CL, Miller DA, Leff T, and Breslow JL (1987). The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron. J Biol Chem 262: 4787-4793.

Hoffer MJV, Hofker MH, van Eck MM, Havekes LM, and Frants RR (1992). Evolutionary conservation of the mouse apolipoprotein e-c1-c2 gene cluster: structure and genetic variability in inbred mice. Genomics 15: 62-67

Hoffer MJV, van Eck MM, Havekes LM, Hofker MH, and Frants RR (1993b). Structure and expression of the mouse apolipoprotein C2 gene. Genomics 17: 45-51.

Hoffer MJV, van Eck MM, Havekes LM, Hofker MH, and Frants RR (1993c). The mouse apolipoprotein C1 gene: structure and expression. Genomics: in press.

Horii A, Nakatsuru S, Ichii S, Nagase H, and Nakamura Y (1993). Multiple forms of the APC gene transcripts and their tissue-specific expression. Hum Mol Genet 2: 283-287.

Horiuchi K, Tajima S, Menju M, and Yamamoto A (1989). Structure and expression of mouse apolipoprotein E gene. J Biochem 106: 98-103.

Karathanasis SK (1985). Apolipoprotein multigene family: tandem organization of human apolipoprotein AI, CIII and AIV genes. Proc Natl Acad Sci USA 82: 6374-6378.

Krumlauf R (1991). Northern blot analysis of gene expression. In: Gene transfer and expression protocols (Murray EJ ed.). Humana Press, Clifton, New Yersey p: 307-360.

Lackner KJ, Law SW, Brewer HB, Sakaguchi AY, and Naylor SL (1984). The human apolipoprotein A-II gene is located on chromosome 1. Biochem Biophys Res Comm 122: 877-883.

Li W-H, Tanimura M, Luo C-C, Datta S, and Chan L (1988). The apolipoprotein multigene family: biosynthesis, structure-function relationships, and evolution. J Lipid Res 29: 245-271.

Luo CC, Li WH, Moore MN, and Chan L (1986). Structure and evolution of the apolipoprotein multigene family. J Mol Biol 187: 325-340.

Lusis AJ, Heinzmann C, Sparks RS, Scott J, Knott TJ, Geller R, Sparkes MC, and Mohandas T (1985). Regional mapping of human chromosome 19: organization of genes for plasma lipid transport (APOC1, -C2, and -E and LDLR) and the genes C3, PEPD and GPI Proc. Natl Acad Sci USA 83: 3929-3933.

Lusis AJ, and Sparks RC (1989). Chromosomal organization of genes involved in plasma lipoprotein metabolism: human and mouse "fat maps". In: Genetic factors in atherosclerosis: approaches and model systems (Lusis AJ, and Sparkes SR, Eds.). Monogr Hum Genet, Basel Karger. vol 12, pp. 79-94.

Shen P, and Howlett GJ (1992). Two coding regions closely linked to the rat apolipoprotein E gene: nucleotide sequences of rat apolipoprotein C-1 and ECL cDNA. Arch Biochem Biophys 297: 345-353.

Whitted BE, Castle CK, Polites HG, Melchior GW, and Marotti KR (1989). Purification, cloning and nucleotide sequence determination of cynomolgus monkey apolipoprotein C-II: comparison to the human sequence. Mol Cell Biochem 90: 69-79.

CHAPTER 6

THE MOUSE LOW DENSITY LIPOPROTEIN RECEPTOR GENE: cDNA SEQUENCE AND EXON-INTRON STRUCTURE.

M.J.V. Hoffer, M.M. van Eck, F. Petrij, A. van der Zee, E. de Wit, D. Meijer, G.Grosveld, L.M. Havekes, M.H. Hofker, and R.R. Frants.

Biochem. Biophys. Res. Commun. 1993, 191: 880-886

THE MOUSE LOW DENSITY LIPOPROTEIN RECEPTOR GENE: cDNA SEQUENCE AND EXON-INTRON STRUCTURE[1]

Mariëtte J.V. Hoffer, Marga M. van Eck, Fred Petrij, André van der Zee, Elly de Wit#, Dies Meijer*, Gerard Grosveld*, Louis M. Havekes#, Marten H. Hofker[2] and Rune R. Frants

MGC-Department of Human Genetics, Leiden University, P.O.Box 9503, 2300 RA Leiden, The Netherlands

#TVVO-TNO, Gaubius Laboratory, P.O.Box 430, 2300 AK Leiden, The Netherlands

*Department of Cell Biology, Erasmus University, P.O.Box 1738, 3000 DR Rotterdam, The Netherlands

Received January 19, 1993

SUMMARY: The low density lipoprotein (LDL) receptor plays a central role in the cholesterol metabolism. The cDNA sequence of the mouse low density lipoprotein receptor (*Ldlr*) gene has been determined and shows 76% homology with the human gene. The exon-intron structure has been determined for the 129/J mouse strain. The gene is composed of 18 exons and spans a region of 28 kb. In addition, the promoter regions of the mouse and human genes are homologous. Northern blot analysis revealed an mRNA of approximately 5 kb. The cloning of the *Ldlr* gene will enhance the usefulness of the mouse for the study of cholesterol metabolism and, in particular, for carrying out gene targeting experiments.

The low density lipoprotein (LDL) receptor mediates the clearance of lipoproteins containing apolipoprotein E (apoE) and apoB-100 from plasma. This receptor-mediated pathway supplies cholesterol to cells in the liver and other organs (1). Mutations in the low density lipoprotein receptor (LDLR) gene are associated with familial hypercholesterolemia characterized by elevated LDL cholesterol levels in the plasma, which lead to an increased risk for atherosclerosis (2). Several animal models like rabbit and monkey (3, 4) have been used to study the role of the LDL receptor in the cholesterol metabolism. Mice have proven to be an important resource for the development of experimental animal models (5). In particular,

[1]Sequence data from this article have been deposited with the EMBL/GenBank/DDBJ Data Libraries under Accession Nos. Z19520 and Z19521.

[2] To whom correspondence should be addressed.

Abbreviations: LDLR: Low Density Lipoprotein Receptor; LDL: Low Density Lipoprotein; APO: Apolipoprotein; RT-PCR: Reverse Transcriptase Polymerase Chain Reaction.

0006-291X/93 $4.00

a powerful approach for the generation of mouse models takes advantage of homologous recombination in pluripotent embryonic stem cells (6) enabling the generation of animals with specific genetic changes.

The use of mice as an animal model for studies on the cholesterol metabolism requires detailed knowledge of the underlying genes. The sequence of the LDLR gene has been reported for several species: human, bovine, rabbit, rat, Xenopus and hamster (7-12). At the genomic level, only the human and hamster LDLR genes have been characterized (12, 13). In the current study we present the coding sequence of the mouse *Ldlr* gene, as well as the exon-intron structure.

MATERIAL AND METHODS

Isolation and sequencing of cDNA clones: A cDNA library derived from liver mRNA of an adult male BALB/c mouse (Clonetech) was screened with a probe derived from the rat LDLR cDNA clone pLDL-1 (10) containing a 2 kb *Eco*RI insert with exons 5-18, and a probe derived from the human LDLR cDNA clone pLDLR-3 (ATCC, # 57005) containing a 1 kb *Pst*I insert with exons 2-8. Phage inserts were isolated by PCR amplification (14) using two PCR primers flanking the *Eco*RI cloning site of λgt11 (5'-GGTGGCGACGACTCCTGGAG-3'; 5'-CGTAATGGTAGCGACCGGCG-3'). After amplification, the PCR products were cloned into the M13mp18/19 vector and sequenced using the T7 sequencing kit (Pharmacia). For RT-PCR analysis, first strand cDNA was synthesized using a primer in exon 12 (5'-CTCAAAGACGGCCAAGGAGA-3'), and AMV reverse transcriptase according to the protocol of the manufacturer (Promega), followed by PCR using primers in exon 1 (5'-TCCTGGCTGCTGCTGGAGTT-3') and in exon 5 (5'-CGAGCTCGTCGCTCATGTCC-3'). The primers in exons 1 and 12 were based on the rat and human LDLR sequence, respectively (7, 10).

Determination of the exon-intron structure: A library was constructed by cloning partial *Mbo*I digests of genomic DNA from the mouse inbred strain 129/J into the *Bam*HI site of the phage EMBL3. This library was screened using a human LDLR cDNA probe (see above), and a mouse cDNA probe containing exons 15-18 (i.e. the *Sall-Eco*RI fragment of mLDLRc8). Phage 1-1 was derived from the Swiss mouse strain and was provided by Dr. A.J. Lusis (University of California, Los Angeles, CA). Individual exons were mapped by restriction enzyme analysis using multiple digests and Southern blot analysis (15) using probes for the individual exons. These probes were obtained from the human LDLR gene by PCR (16). Some intron sizes were established by PCR analysis using a primer in exon 18 (5'-TGCCACATCATCCTCCAGGC-3'), and in exon 16 (5'-ACATGGCTGGCAGAGGGAAT-3') or 17 (5'-AGGAACTGGCGGCTGAGGAA-3'). The primers in exons 17 and 18 were based on the rat cDNA sequence (10).

Northern blotting and ligand blotting: Poly(A)$^+$ RNA was isolated from liver using the polyATract mRNA isolation kit (Promega) and Northern blot analysis was performed according to standard protocols (17) using mLDLRc8 as a probe. Membrane associated proteins were isolated from a mouse liver, and used for ligand blotting as described by Lombardi *et al.* (18) using human ^{125}I-LDL.

RESULTS AND DISCUSSION

Isolation of the cDNA encoding the mouse LDL receptor. A mouse liver cDNA library was screened with a rat LDLR cDNA probe. Clone mLDLRc8 was obtained which is 1.5

```
     M  S  T  A  D  L  M  R  R  W  V  I  A  L  L  L  A  A  A  G  V  A  V  E  D  S  G  S
   1 ATGAGCACCGCGGATCTGATGCGTCGCTGGGTCATCGCCCTGCTCCTTGCTGCTGCCGGAGTTGCAGTAGAAGACTCAGGCAGC
     R  N  E  F  Q  C  R  D  G  K  C  I  A  S  K  W  V  C  D  G  S  P  E  C  P  D  G  S
  85 AGGAACGAGTTCCAGTGTAGAGACGGAAAATGCATCGCTAGCAAGTGGGTGTGCGATGGCAGCCCCGAGTGCCCGGATGGCTCC
     D  E  S  P  K  T  C  M  S  V  T  C  Q  S  N  Q  F  S  C  G  G  R  V  S  R  C  I  P
 169 GATGAGTCCCCAAAGACATGCATGTCTGTCACCTGTCAGTCCAATCAATTCAGCTGTGGAGGCCGTGTCAGCCGATGCATTCCT
     D  S  W  R  C  D  G  Q  V  D  C  E  N  D  S  D  E  Q  G  C  P  P  K  T  C  S  Q  D
 253 GACTCCTGGAGATGTGATGGACAGGTAGACTGTGAAAATGACTCAGACGAACAAGGCTGTCCCCCCAAGACGTGCTCCCAGGAT
     D  F  R  C  Q  D  G  K  C  I  S  P  Q  F  V  C  D  G  D  R  D  C  L  D  G  S  D  E
 337 GACTTCCGATGCCAGGATGGCAAGTGCATCTCCCCGCAGTTTGTGTGTGATGGAGACCGAGATTGCCTAGATGGCTCTGATGAG
     A  H  C  P  A  T  T  C  G  P  A  H  F  R  C  K  S  S  I  C  I  P  S  L  W  A  C  D
 421 GCCCACTGCCCAGCCACCACTTGTGGCCCCGCCCACTTCCGCTGCAAATCATCCATATGCATCCCCAGTCTTTGGGCCTGCGAC
     G  D  V  D  C  V  D  G  S  H  E  W  P  Q  N  C  Q  A  E  D  T  A  S  K  G  V  S  S
 505 GGGGATGTCGACTGTGTTGACGGCTCCCATGAGTGGCCACAGAACTGCCAGGCCGAAGACACGGCCTCCAAAGGCGTTAGCAGC
     P  C  S  S  L  E  F  H  C  G  S  S  E  C  I  H  R  S  W  V  C  D  G  E  A  D  C  K
 589 CCCTGCTCCTCCCTGGAGTTCCACTGTGGTAGCAGTGAGTGTATCCATCGCAGCTGGGTCTGTGACGGCGAGGCAGACTGCAAG
     D  K  S  D  E  E  H  C  A  V  A  T  C  R  P  D  E  F  Q  C  A  D  G  S  C  I  H  G
 673 GACAAGTCAGATGAGGAGCACTGCGCGGTGGCCACCTGCCGACCTGATGAATTCCAGTGTGCAGATGGCTCCTGCATTCACGGT
     S  R  Q  C  D  R  E  H  D  C  K  D  M  S  D  E  L  G  C  V  N  V  T  Q  C  D  G  P
 757 AGCCGCCAGTGTGACCGTGAACATGACTGCAAGGACATGAGCGACGAGCTCGGCTGCGTCAATGTGACACAGTGTGATGGCCCC
     N  K  F  K  C  H  S  G  E  C  I  S  L  D  K  V  C  D  S  A  R  D  C  Q  D  W  S  D
 841 AACAAGTTCAAGTGTCACAGTGGGGAGTGCATCAGCTTGGACAAGGTGTGCGACTCCGCCCGCGACTGCCAGGACTGGTCGGAT
     E  P  I  K  E  C  K  T  N  E  C  L  D  N  N  G  G  C  S  H  I  C  K  D  L  K  I  G
 925 GAGCCCATCAAGGAGTGCAAGACCAACGAGTGTTTGGACAACAATGGTGGCTGTTCCCACATCTGCAAGGACCTCAAGATTGGC
     S  E  C  L  C  P  S  G  F  R  L  V  D  L  H  R  C  E  D  I  D  E  C  Q  E  P  D  T
1009 TCTGAGTGCCTGTGTCCCAGCGGCTTCCGGTTGGTGGACCTCCACAGGTGTGAAGATATTGACGAGTGTCAGGAGCCAGACACC
     C  S  Q  L  C  V  N  L  E  G  S  Y  K  C  E  C  Q  A  G  F  H  M  D  P  H  T  R  V
1093 TGCAGCCAGCTCTGTGTGAACCTGGAAGGCAGCTACAAGTGTGAGTGCCAGGCCGGCTTCCACATGGACCCACACACCAGGGTC
     C  K  A  V  G  S  I  G  Y  L  L  F  T  N  R  H  E  V  R  K  M  T  L  D  R  S  E  Y
1177 TGCAAGGCTGTGGGCTCCATAGGCTATCTGCTCTTCACCAACCGCCACGAGGTCCGGAAGATGACCCTGGACCGCAGCGAGTAC
     T  S  L  L  P  N  L  K  N  V  V  A  L  D  T  E  V  T  N  N  R  I  Y  W  S  D  L  S
1261 ACCAGTCTGCTCCCCAACCTGAAGAATGTGGTGGCTCTCGACACGGAGGTGACCAACAATAGAATCTACTGGTCGGACCTGTCC
     Q  K  K  I  Y  S  A  L  M  D  Q  A  P  N  L  S  Y  D  T  I  I  S  E  D  L  H  A  P
1345 CAAAAAAAGATCTACAGCGCCCTGATGGACCAGGCCCCTAACTTGTCCTACGACACCATCATCAGTGAGGACCTGCATGCCCCT
     D  G  L  A  V  D  W  I  H  R  N  I  Y  W  T  D  S  V  P  G  S  V  S  V  A  D  T  K
1429 GACGGGCTGGCGGTAGACTGGATCCACCGCAACATCTACTGGACAGATTCAGTCCCGAGCGTATCTGTGGCTGACACCAAG
     G  V  K  R  R  T  L  F  Q  E  A  G  S  R  P  R  A  I  V  V  D  P  V  H  G  F  M  Y
1513 GGCGTAAAGAGGAGGACACTGTTCCAAGAGGCAGGGTCCAGACCCAGAGCCATCGTAGTGGACCCTGTGCATGGCTTCATGTAC
     W  T  D  W  G  T  P  A  K  I  K  K  G  G  L  N  G  V  D  I  H  S  L  V  T  E  N  I
1597 TGGACAGATTGGGGAACACCCGCCAAGATCAAGAAAGGGGGTTTGAATGGTGTGGACATCCACTCACTGGTGACCGAAAACATC
     Q  W  P  N  G  I  T  L  D  L  S  S  G  R  L  Y  W  V  D  S  K  L  H  S  I  S  S  I
1681 CAGTGGCCAAATGGCATCACACTAGATCTTTCCAGTGGCCGTCTCTATTGGGTTGATTCCAAACTCCACTCTATCTCCAGCATC
     D  V  N  G  G  N  R  K  T  I  L  E  D  E  N  R  L  A  H  P  F  S  L  A  I  Y  E  D
1765 GATGTCAATGGGGGCAATCGGAAAACCATTTTGGAGGATGAGAACCGGCTGGCCCACCCCTTCTCCTTGGCCATCTATGAGGAC
     K  V  Y  W  T  D  V  I  N  E  A  I  F  S  A  N  R  L  T  G  S  D  V  N  L  V  A  E
1849 AAAGTGTATTGGACAGATGTCATAAACGAAGCCATTTTCAGTGCCAATCGACTCACGGGTTCAGATGTGAATTTGGTGGCTGAA
     N  L  L  S  P  E  D  I  V  L  F  H  K  V  T  Q  P  R  G  V  N  W  C  E  T  T  A  L
1933 AACCTCTTGTCCCCGGAGGACATTGTCCTGTTCCACAAGGTCACACAGCCTAGAGGGGTGAACTGGTGTGAGACAACAGCCCTC
     L  P  N  G  G  C  Q  Y  L  C  L  P  A  P  Q  I  G  P  H  S  P  K  F  T  C  A  C  P
2017 CTCCCCAATGGTGGTTGCCAGTACCTGTGCCTGCCCGCCCCACAGATCGGTCCCCACTCGCCCAAATTCACCTGCGCCTGCCCT
     D  G  M  L  L  A  K  D  M  R  S  C  L  T  E  V  D  T  V  L  T  T  Q  G  T  S  A  V
2101 GATGGCATGCTGCTGGCCAAGGACATGCGGAGCTGCCTCACAGAAGTCGACACTGTACTGACCACCCAGGGGACATCCGCCGTC
     R  P  V  V  T  A  S  A  T  R  P  P  K  H  S  E  D  L  S  A  P  S  T  P  R  Q  P  V
2185 CGGCCTGTGGTCACCGCATCAGCTACCAGGCCACCGAAGCACAGTGAGGATCTCTCAGCTCCCAGTACTCCTAGGCAGCCTGTG
     D  T  P  G  L  S  T  V  A  S  V  T  V  S  H  Q  V  Q  G  D  M  A  G  R  G  N  E  E
2269 GACACCCCAGGGCTCAGCACAGTGGCGTCAGTGACAGTGTCCCACCAAGTCCAGGGTGACATGGCTGGCAGAGGGAATGAGGAG
     Q  P  H  G  M  R  F  L  S  I  F  F  P  I  A  L  V  A  L  L  V  L  G  A  V  L  L  W
2353 CAGCCACATGGTATGAGGTTCCTGTCCATCTTCTTCCCTATTGCACTGGTTGCCCTCCTTGTCCTTGGGGCCGTCCTGCTGTGG
     R  N  W  R  L  K  N  I  T  I  N  S  I  N  F  D  N  P  V  Y  Q  K  T  T  E  D  E  L
2437 AGGAACTGGCGGCTGAAGAACATCACAATCAACAGCATAAACTTTGACAACCCAGTCTACCAGAAGACCACAGAGGACGAGCTC
     H  I  C  R  S  Q  D  G  Y  T  Y  P  S  R  Q  M  V  S  L  E  D  D  V  A  *
2521 CACATTTGCCGAAGCCAGGATGGCTATACCTACCCCTCAAGACAGATGGTCAGCCTGGAGGACGATGTGGCATGAGCAGCCGGG
2605 AGAGCCGTCTCTTTCCGGGATCCATTGCCAAGCTTAGGCAGAAAAGACACTCTCTCCAGACCTCCCCATCCAGCACTGGTCCTG
2689 CCACCTCCCTGGGNTCTGTGTTGCTCAAAGCAAGATAGAGCAAAGCTGGGCTGGGGGGCCAAGCTCAGCTTCCTGTCTGCCCCA
2773 GGTTCTGTTTTATATATTTATTGTCTGGGACAGAAAAGGCTACTGGCTGTGCTTGAAATTCGA
```

Fig. 1. Nucleotide sequence and deduced amino acid sequence of the mouse *Ldlr* cDNA. The deduced amino acid sequence is indicated above and nucleotides are numbered at the left of the nucleotide sequence; nucleotide 1 is the A of the putative translation initiation ATG codon. The putative stop codon is indicated by an asterisk.

kb in size, and contains exons 8-17 and a part of exon 18. Using a human LDLR cDNA probe, clone mLDLRc44 was obtained which is 1.0 kb in size and contains exons 4-10. The 5'-end of the gene was obtained by RT-PCR using primers in exons 1 and 5. A fragment of 700 bp was isolated and cloned (mLDLRc90). These 3 cDNA clones were mapped with several restriction enzymes and subcloned for sequence analysis. The cDNA sequence was

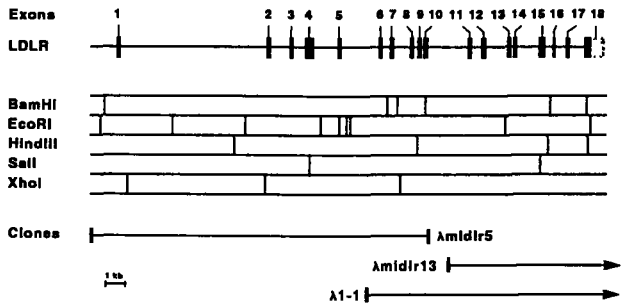

Fig. 2. Exon-intron structure of the mouse *Ldlr* gene. The restriction enzyme map for the enzymes indicated at the left is shown for the inbred strain 129/J. The exons are indicated with solid boxes. Stippled box indicates a part of exon 18 which was not characterized. The phage clones used for restriction mapping are indicated below the map.

determined from both strands. The sequence of exon 1 was determined at the genomic level (described below). Fig. 1 shows the nucleotide sequence of the entire coding region and a small portion of the 3'-untranslated region of mouse *Ldlr* cDNA with the deduced amino acid sequence. The cDNA encodes a protein of 864 amino acid residues, including a signal peptide of 21 amino acid residues. The mouse *Ldlr* sequence shows an overall homology of 76% with the human LDLR sequence (11). Several domains of the LDLR known to be important for its function show a higher conservation, i.e. the EGF precursor domain (83%), and the cytoplasmic domain (85%). LDLR genes of other species show a similar pattern of conservation (8-12).

Exon-intron structure of the mouse *Ldlr* gene. A genomic library, derived from DNA of the 129/J mouse strain was screened with a human LDLR cDNA probe and a mouse *Ldlr* cDNA probe. Two genomic clones, λmLDLR5 and λmLDLR13, were obtained spanning nearly the entire *Ldlr* gene, 1.5 kb of the 5'-flanking region, and approximately 12 kb of the 3'-flanking region. Another clone (λ1-1) derived from the Swiss mouse strain spans the small region between exon 10 and 11 which was not contained within λmLDLR5 or λmLDLR13 (Fig. 2). A restriction map of the *Ldlr* gene was constructed using 5 different enzymes (Fig. 2). Phage DNA was digested with these enzymes and additional restriction enzymes recognizing sites within particular exons. Single and multiple digests were analyzed by Southern blotting using exon-specific probes. Due to the generation of short restriction fragments (generally between 200 and 500 bp) the position of the exons could be determined with high resolution. The location of exons which could not be determined by restriction mapping were established by subcloning and sequence analysis (exons 1, 11 and 13). In addition, PCR analysis using forward primers in exons 16 or 17, and a backward primer in

```
Mouse  gaaattctgtgggaggaatttgaggaacttcccactgctgcgggagcttctggggttaaaagagacgatgtcacatcggc -258
       :     :   : ::    :: :: : :::       :   :  :::::::  ::::::::::   ::::::::::::::::
Human  gcctgccctggcgacactttcgaaggactggagtgggaatc-agagcttcacgggttaaaag--ccgatgtcacatcggc -203
                repeat 1                                  repeat 2       repeat 3
Mouse  cgttccaagctcctcccagctcagtgaggtgaagatttttgaaaatcaccccattgcagactcctccccggcctggaaac -178
       ::::: :: :::::::     ::::::::::::::  :::::::::::::::: :::: ::::::::::   :: :::::
Human  cgttcgaaactcctcctcttgcagtgaggtgaagacatttgaaaatcaccccactgcaaactcctccccctgctagaaac -123

Mouse  ctcgc-ccctagtactgggaatgactctgggcgtgcggcgtagtttgcagccgggacaccgtgaggcttgcgagcccaga -99
       ::: :    : :    : ::          :::: :::   :  ::::  ::
Human  ctcacattgaaatg-ctgtaaatga------cgtg-ggccccgagtgcaatcgc---------------------------- -77

Mouse  ttcgcagccgagacaccgtggggcccgcgatccagtgtttgcagcgggaacatttcggggtctgtgatccgagtgaggac -19
                               :  : :::: :::: ::::  : :::   :  :::: :::::::: ::  ::::
Human  ------------------------ggga-agccagggtttccagctaggacacagc-aggtc-gtgatccgggtcgggac -24

Mouse  ---gcaacgcagaagct--aaggATGAGCACCGCGGATCTGATGCGTCGCTGGGTCATCGCCCTGCTCCTTGCTGCTGCC  58
          ::   ::::: :::   :: ::: :  ::  ::    ::     ::::::  : ::::: ::::::: :: :: ::
Human  actgcctggcagaggctgcgagcATGGGGCCCTGGGGCTGGAAATTGCGCTGGACCGTCGCCTTGCTCCTCGCCGCGGCG  58

Mouse  GGAGTTGCAGgtaaggcttgcgcca  67
       ::   ::::::::::::::::: :::
Human  GGGACTGCAGgtaaggcttgctcca  67
```

Fig. 3. Comparison of the 5'-flanking region and exon 1 of the mouse and human LDLR genes. Identical nucleotides are marked by (:). Gaps (-) have been inserted to achieve maximum homology. The direct repeats are indicated by arrows. Exon 1 is indicated by upper case. Human data are according to Südhoff *et al.* (13).

exon 18 provided the genomic distance between these exons. The mouse *Ldlr* gene contains 18 exons and spans a region of 28 kb. The overall exon-intron structure agrees with the structure described for the corresponding hamster and human genes (7, 12). Comparison of the LDLR gene between man and mouse reveals that the introns in the mouse gene are generally smaller in size. The restriction maps of the *Ldlr* gene of the 129/J inbred mouse strain and the Swiss mouse are similar. However, an insertion of approximately 500 bp was found in intron 12 for the inbred strain 129/J. Two additional sites were found for the restriction enzymes *Eco*RI and *Hin*dIII in introns 5 and 6 respectively for the Swiss mouse (data not shown). These three sites correspond to restriction fragment length variations, which were described previously (19) using Southern blot analysis and indicate genetic variation within the *Ldlr* gene between different inbred mouse strains.

Expression of the mouse *Ldlr* gene. The nucleotide sequence of the 5'-flanking region of the mouse and the human genes are highly homologous (Fig. 3). Three direct repeats, involved in transcriptional activation and regulation by sterols in the human LDLR gene (13), were also present in the mouse promoter region. Northern blot analysis using mouse liver poly(A)$^+$ mRNA demonstrated a major transcript of approximately 5 kb in size (Fig. 4A). The open reading frame consists of 2592 nucleotides. These results predict a 3'-untranslated region of approximately 2.4 kb. Ligand blotting reveals, similar to the situation of the human gene, a precursor with an apparent molecular weight of 116 kD and a mature protein of 135 kD (Fig. 4B).

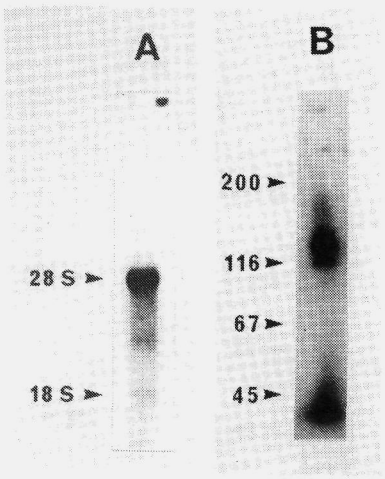

Fig. 4. (A) Northern blot analysis of mouse liver RNA. Northern blot of poly(A)$^+$ RNA (10 μg) from liver was hybridized using the cDNA clone mLDLRc8 as a probe. The 28S and 18S ribosomal RNA bands are indicated at the left. (B) Ligand blotting of the mouse LDL receptor. Mouse liver was used to isolate membrane bound proteins and used for ligand blotting with human ^{125}I-LDL according to Lombardi *et al.* (18). The molecular weight is given in kD. A degradation product can be observed at the low molecular weight range.

Due to the rapid progress of transgenic technology, mice have become an extremely valuable tool for the study of cholesterol metabolism. The LDL receptor plays a central role in this metabolism, emphasising the need for the isolation and characterization of the LDLR gene for future genetic studies. The exon-intron structure has been determined for the 129/J mouse strain. This strain is commonly used for gene targeting via homologous recombination which is a critical step in this process. Therefore the 129/J clones described in this study could be particularly useful, since it has recently been shown that recombination frequencies are higher using isogenic DNA (20, 21).

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr A.J. Lusis for the LDLR phage and Squibb for the rat LDLR cDNA probe. This work was supported by HGO-TNO (#900-539-052) and the Royal Netherlands Academy of Arts and Sciences (M.H.H.).

REFERENCES

1. Brown, M.S., and Goldstein, J.L. (1986) Science 232, 34-47.
2. Goldstein, J.L. and Brown, M.S. (1989) in The Metabolic Basis of Inherited Disease 6th Edn. (Scriver, C.R., Beaudet, A., Sly, W.S., and Valle, D., eds.), pp. 1215-1250, McGraw-Hill, New York.

3. Watanabe, Y. (1980) Atherosclerosis 36, 261-268.
4. Scanu, A.M., Khalil, A., Neven, L., Tidore, M., Dawson, G., Pfaffinger, D., Jackson, E., Carey, K.D., McGill, H.C., and Fless, G.M. (1988) J. Lipid. Res. 29, 1671-1681.
5. Hofmann, S.L., Russell, D.W., Brown, M.S., Goldstein, J.L., and Hammer, R.E. (1988) Science 239, 1277-1281.
6. Capecchi, M.R. (1989) Trends Genet. 5, 70-76.
7. Yamamoto, T., Davis, C.G., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L., and Russell, D.W. (1984) Cell 39, 27-38.
8. Russell, D.W., Yamamoto, T., Schneider, W.J., Slaugther, C.J., Brown, M.S., and Goldstein, J.L. (1983) Proc. Natl. Acad. Sci. USA 80, 7501-7505.
9. Yamamoto, T., Bishop, R.W., Brown, M.S., Goldstein, J.L., and Russell, D.W. (1986) Science 232, 1231-1237.
10. Lee, L.Y., Mohler, W.A., Schafer, B.L., Freudenberger, J.S., Byrne-Connolly, N., Eager, K.B., Mosley, S.T., Leigthon, J.K., Thrift, R.N., Davis, R.A., and Tanaka, R.D. (1989) Nucleic Acids Res. 17, 1259-1260.
11. Metha, K.D., Chen, W-J., Goldstein, J.L., and Brown, M.S. (1991) J. Biol. Chem. 266, 10406-10414.
12. Bishop, R.W. (1992) J. Lipid Res. 33, 549-557.
13. Südhoff, T.C., Goldstein, J.L., Brown, M.S., and Russell, D.W. (1985) Science 228, 815-822.
14. Hoffer, M.J.V., Hofker, M.H., van Eck, M.M., Havekes, L.M., and Frants, R.R. (1993). Genomics, in press.
15. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, rev. ed., vols 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
16. Leitersdorf, E., Tobin, E.J., Davignon, J., and Hobbs, H.H. (1990) J. Clin. Invest. 85, 1014-1023.
17. Krumlauf, R. (1991) in Gene Transfer and Expression protocols (Murray, E.J., ed.), pp. 307-360, Humana Press, Clifton, New Jersey.
18. Lombardi, P., de Witt, E., Frants, R.R., and Havekes, L.M. (1990) Biochim. Biophys. Acta 1044, 127-132.
19. Srivastava, R.A.K., Jiao, S., Tang, J., Pfleger, B.A., Kitchens, R.T., and Schonfeld, G. (1991) Biochim. Biophys. Acta 1086, 29-43.
20. te Riele, H., Robanus Maandag, E., and Berns, A. (1992) Proc. Natl. Acad. Sci. USA 89, 5128-5132.
21. van Deursen, J. and Wieringa, B. (1992) Nucleic Acids Res. 20, 3815-3820.

SUMMARY & SAMENVATTING

SUMMARY

Heterogeneity in environmental and genetic background in man leads to a variability in plasma levels of lipids and lipoproteins. Since variations in lipoprotein levels are strongly associated with susceptibility to atherosclerosis, the characterization of the hereditary influences provides an opportunity to a better understanding of the genetic components involved in this major human disease. The mouse provides an animal model for examining the heritable variation of lipoproteins. Many inbred strains are available and form a source of infinite numbers of genetically identical individuals. Together with a well defined genetic map, the use of mouse inbred strains will contribute to the identification of genes controlling the variability of lipoprotein levels. In addition, new technologies, like transgenesis and gene targeting, have been developed for manipulating genes of interest. However, despite the increased popularity of mouse models at present, still very little is known about the genes involved in lipoprotein metabolism in the mouse. The aim of this study was to increase our knowledge about these relevant mouse genes. In this thesis, we have focused on the *Apoe-c1-c2* gene cluster and the *Ldlr* gene which both play an important role in the metabolism of atherogenic lipoproteins.

In order to characterize the *Apoe-c1-c2* gene cluster, overlapping cosmid clones were obtained (Chapter 2). We have demonstrated that the *Apoe-c1-c2* gene cluster is evolutionarily conserved in the mouse on the basis of Southern blot analysis. In line with the hypothesis that the duplication of the APOC1 gene took place after the divergence of rodents and primates, the mouse lacks the equivalent of a human pseudo APOC1 gene. We have also determined the genetic variation within this gene cluster. A variable number of tandem repeats (VNTR), in the third intron of human APOC2 was found to be located at an identical position in the mouse gene. Despite a high copy number, this VNTR displayed only two variants among several inbred mouse strains. With the aid of six restriction fragment length variants (RFLVs) only two haplotypes could be deduced among the 11 different inbred strains tested. The absence of genetic variation in this gene cluster could be a result of selection during the generation of inbred mouse strains.

The *Apoc1* gene (Chapter 3) and the *Apoc2* gene (Chapter 4) were characterized in more detail. The *Apoc1* gene, located 3.4 kb 3' of the *Apoe* gene, was found to span a region of approximately 3.3 kb and consists of 4 exons. The exon-intron structure

resembled the overall gene structure of the human APOC1 gene with the exception of introns 2 and 3 which were considerably smaller. This is probably due to the lower number of *Alu*-like repeated elements in the mouse gene. The transcription start site has been determined using primer extension analysis. The mouse *Apoc1* promoter region shows 80% homology with the human counterpart. Interestingly, a 15 bp sequence motif showed 90% homology with a regulatory element in the human APOE gene. The *Apoc1* gene is primarily expressed in the liver by a transcript of 0.4 kb in size encoding a protein of 88 amino acids. Comparisons of this amino acid sequence with apoC1 sequences of other species revealed discrete regions with a high degree of conservation.

The mouse *Apoc2* mxne consists of four exons spanning a region of approximately 6 kb (Chapter 4). Transcripts of 0.5 kb in size were found in fetal liver, adult liver, intestine and peritoneal macrophages. The *Apoc2* transcript encodes a protein of 97 amino acids which, like in the rat apoC2, is one residue shorter.

Interestingly, in addition to the normal sized cDNA clones, an unusual large cDNA clone was found containing additional sequences at the 5'-end. The additional 5' sequences were encoded by two exons located 5' of the liver start site of *Apoc2*. Evolutionary conservation between man and mouse of the 5' additional sequences was confirmed by Southern blot analysis. The expression level of this large transcript in the tissues used for Northern blot analysis was below the detection level. Therefore, the more sensitive reverse transcriptase (RT) PCR analysis was used and transcripts were found indicating alternative splicing.

Chapter 5 describes the isolation of cDNA clones encoding a novel gene within the *Apoe-c1-c2* gene cluster of the mouse using a probe containing only the 5' additional sequences of the large *Apoc2* cDNA clone. This gene, designated apolipoprotein C2 linked (*Acl*) gene was found to span a region of 3.6 kb and consists of three exons located 5' of the *Apoc2* gene. The *Acl* gene is highly expressed in the liver with a transcript of approximately 0,5 kb in size and encodes a putative protein of 124 amino acid residues.

The finding of a novel gene 5' of the *Apoc2* gene implies that the transcripts containing both genes must have been arisen from "read through" events beyond the polyadenylation signal of the 5' *Acl* gene. Two previously reported cDNA sequences from other species are homologous to this novel gene. From cynomolgus monkey, the first 250 bp of an APOC2 cDNA clone show a high homology with the mouse *Acl* gene and resembles the fusion transcript represented by the mAPOC2c4 cDNA clone

of mouse *Apoc2*. In addition, homology was found with a cDNA clone which could represent an additional gene within the APOE-C1-C2 gene cluster of the rat. Comparisons of the *Acl* gene with the rat ECL sequences showed a high homology. However, the orientation of the open reading frame implied that the transcription would be in the opposite direction in the rat.

The function of this novel gene within the *Apoe-c1-c2* gene cluster remains still unclear. Searches in data banks did not reveal recognizable motifs sofar. Further studies with the *Acl* protein indicated that *Acl* shows resemblance with other apolipoproteins. However, further studies are needed to determine whether its function also resembles that of apolipoproteins.

The final part of this thesis focuses on the *Ldlr* gene, which also plays an important role in the metabolism of atherogenic lipoproteins. The cDNA sequence containing the coding region and a small part of the 3' untranslated region encodes a protein of 864 amino acid residues, including a signal peptide of 21 amino acid residues. Comparison of the mouse *Ldlr* sequences with the human LDLR gene showed a high conservation, especially within the domains which are important for its function. In addition, Northern blot analysis and ligand blotting showed that the mouse *Ldlr* mRNA and protein are similar to the human counterpart.

The exon-intron structure of the *Ldlr* gene was determined using Southern blot analysis to map the individually exons. The mouse *Ldlr* gene consists of 18 exons, spanning a region of 28 kb. Comparison of the LDLR gene between man and mouse revealed that the introns in the mouse gene were generally smaller in size. The regulatory elements in the promoter region of the human LDLR gene were also present in the mouse gene and are highly homologous.

The delineation of the sequence and structural organization of the genes described in this thesis will enable us to improve the usefulness of mouse models for studying the genetic aspects of lipoprotein metabolism. The results found for the mouse *Apoc1*, *Apoc2*, and *Ldlr* genes indicate a strong conservation of these genes on DNA and protein level. This resemblance between human and mouse genes makes the mouse a useful animal model. Furthermore, the delineation of the genomic structure of the genes described in this thesis will enable us to carry out gene targeting experiments. These experiments will be very useful for unraveling the physiological function of the corresponding proteins. Also, information about the evolutionary conservation of these genes may provide information about regulatory sequences within the genes or functional domains of the proteins.

SAMENVATTING

Bij de mens leidt heterogeniteit in omgevingsfactoren en genetische achtergrond tot variatie in plasma lipiden en lipoproteïnen concentraties. Variatie in lipoproteïnen concentraties is sterk geassocieerd met de aanleg voor atherosclerose. Karakterisatie van de erfelijke factoren die deze variatie beïnvloeden geeft de mogelijkheid tot een beter inzicht in de genetische componenten betrokken bij hart- en vaatziekten. Voor het bestuderen van deze genetische componenten is de muis een geschikt proefdier model. Een groot aantal inteelt stammen zijn beschikbaar en vormen een bron van onuitputtelijke aantallen van genetisch identieke individuen. De genetische kaart is uitgebreid beschreven en zorgt er voor dat de muis een aantrekkelijk proefdier model is voor het identificeren van genen betrokken bij de variatie in lipoproteïnen. Daarnaast zijn er nieuwe technieken beschikbaar voor het manipuleren van kandidaat genen in de muis. Voldoende kennis van de betrokken genen in de muis is belangrijk voor de juiste interpretatie van resultaten verkregen met deze studies. Ondanks een toename in de populariteit van de muis als proefdier model is er nog maar weinig bekend over de muizen genen die belangrijk zijn voor het lipoproteïnen metabolisme. Dit proefschrift beschrijft een aantal relevante muizen genen om dit hiaat in onze kennis te verkleinen. Deze studie beperkt zich tot het *Apoe-c1-c2* gencluster en het *Ldlr* gen die beide een belangrijke rol spelen in het metabolisme van atherogene lipoproteïnen.

Overlappende cosmid kloons werden geïsoleerd om het *Apoe-c1-c2* gencluster te karakteriseren (Hoofdstuk 2). Met behulp van Southern blot analyse hebben we aangetoond dat het *Apoe-c1-c2* gencluster evolutionair geconserveerd is in de muis. De muis mist het equivalent van het humane pseudo APOC1 gen wat in overeenstemming is met de hypothese dat de duplicatie van het APOC1 gen plaats vond na de afsplitsing van knaagdieren en primaten. Daarnaast hebben we de genetische variatie van dit gencluster bepaald. Een variabel aantal tandem repeats (VNTR), in het derde intron van het humane APOC2 gen, werd op dezelfde positie gelokaliseerd in het muizen gen. Ondanks het hoge kopie aantal werden er maar twee varianten gevonden bij verschillende inteelt stammen. Naast de VNTR werd ook voor zes restrictie fragment lengte polymorfismen (RFLPs) de genetisch variatie bepaald. Maar twee verschillende haplotypen werden gevonden bij de 11 geteste inteelt stammen. De afwezigheid van genetisch variatie in dit gencluster kan veroorzaakt zijn door selectie gedurende de generatie van deze inteelt stammen.

Hoofdstuk 3 en 4 beschrijven de karakterisatie van respectievelijk het *Apoc1* gen en het *Apoc2* gen. Het *Apoc1* gen, 3,4 kb 3' van het *Apoe* gen gelegen, is 3,3 kb groot en bestaat uit 4 exonen. De exon-intron structuur vertoond grote overeenkomsten met het humane gen met uitzondering van intron 2 en 3 die beiden aanzienlijk kleiner zijn. Dit wordt waarschijnlijk veroorzaakt door een kleiner aantal *Alu*-repeats in het muizen gen. Het transcriptie startplaats is bepaald met behulp van primer extentie analyse. De promotor gebied van muizen *Apoc1* vertoond 80% homologie met het overeenkomstige humane gen. Interessant was dat een 15 bp motief 90% homologie vertoonde met een regulatie element in het humane APOE gen. Het *Apoc1* gen komt voornamelijk in de lever tot expressie en het transcript van 0,4 kb codeert voor een eiwit van 88 aminozuren. Vergelijking van de muizen *Apoc1* aminozuur sequentie met die van andere soorten toonde een sterke conservering in bepaalde stukken.

Het muizen *Apoc2* gen bestaat uit 4 exonen en is ongeveer 6 kb groot (Hoofdstuk 4). Een transcript van 0,5 kb in grootte werd aangetoond in foetaal lever, adulte lever, dunne darm en peritoneale macrofagen. Het *Apoc2* transcript codeert voor een eiwit van 97 aminozuren dat, net als in de rat, één residu korter is en de geconserveerde plaats voor proteolyse van het pro-eiwit mist vergeleken met het humane APOC2 gen.

Naast de cDNA kloons van normale grootte werd een ongewoon grote cDNA kloon, mAPOC2c4, met aan de 5' kant additionele sequenties geïsoleerd. Met behulp van Southern blot analyses werd evolutionaire conservering tussen mens en muis aangetoond. Expressie van dit grote transcript kon niet met Northern blot analyse worden aangetoond. Met behulp van de gevoeligere RT-PCR techniek werden transcripten gevonden die ontstaan waren door alternatieve splicing.

Hoofdstuk 5 beschrijft de isolatie van cDNA klonen, met behulp van een probe die alleen de 5' additionele sequenties bevat van het mAPOC2c4 cDNA kloon, die coderen voor een nieuw gen in het *Apoe-c1-c2* gencluster. Dit nieuwe gen, het apolipoproteïne C2 gecorreleerde (*Acl*) gen genaamd, is ongeveer 3,6 kb lang en bestaat uit 3 exonen die 5' van het *Apoc2* gen liggen. Het *Acl* gen komt sterk tot expressie in de lever en het transcript dat 0,5 kb lang is codeert voor een eiwit van 124 aminozuren.

De aanwezigheid van een nieuw gen 5' van het *Apoc2* gen impliceert dat het transcript dat beide genen bevat ontstaan is door het overslaan van het polyadenylerings signaal van het *Acl* gen. Twee eerder gerapporteerde cDNA sequenties zijn homoloog aan dit gen. De eerst 250 bp van een rhesus aap APOC2 cDNA sequentie is sterk homoloog met het muizen *Acl* gen en vertoond een grote gelijkenis met het fusie transcript, de *Apoc2* cDNA kloon mAPOC2c4. Daarnaast werd

er homologie gevonden met een cDNA sequentie waarvan verondersteld werd dat het een nieuw gen vertegenwoordigt in het APOE-C1-c2 gencluster van de rat. De rat ECL sequentie vertoond een sterke homologie met het *Acl* gen op één uitzondering na, de oriëntatie van het open reading frame is tegenovergesteld aan dat van het *Acl* gen. De functie van dit nieuwe gen in het *Apoe-c1-c2* gencluster is tot nu toe onbekend en verdere studies zijn nodig om de functie van dit gen uit te zoeken.

Het laatste deel van het proefschrift beschrijft het *Ldlr* gen dat een belangrijke rol speelt in het metabolisme van atherogene lipoproteïnen. De cDNA sequentie dat het coderende deel en een klein deel van het 3' onvertaalde gebied bevat, codeert voor een eiwit van 864 aminozuren inclusief een signaal peptide van 21 aminozuren. Vergelijking van de muizen *Ldlr* sequentie met het humane gen laat een sterke conservering zien met name in de domeinen die belangrijk zijn voor de functie van het eiwit. Verder vertonen het mRNA en het eiwit grote overeenkomsten met de humane equivalenten.

De exon-intron structuur van het *Ldlr* gen werd bepaald door middel van Southern blot analyse om ieder individueel exon in kaart te brengen. Het muizen *Ldlr* gen bestaat uit 18 exonen en is 28 kb lang. In het muizen gen zijn de intronen relatief kleiner vergeleken met die in het humane gen. De regulatie elementen in het promotor gebied van het humane gen werden ook bij de muis gevonden en waren sterk homoloog.

De resultaten beschreven in dit proefschrift bevorderen de toepassing van de muis als proefdiermodel voor het bestuderen van de genetische aspecten van het lipoproteïnen metabolisme. De resultaten gevonden voor de muizen *Apoc1*, *Apoc2* en *Ldlr* genen laten een sterke conservering zien van deze genen op DNA en eiwit niveau. Door deze gelijkenis tussen humane en muizen genen is de muis een geschikt proefdier model. De karakterisatie van de genomische structuur van de genen beschreven in dit proefschrift maakt het mogelijk om "gene targeting" experimenten op te zetten. Deze experimenten zullen een belangrijke bijdrage gaan leveren aan de bepaling van de functies van de bijbehorende eiwitten. Ook kan informatie over de evolutionaire conservering van deze genen aanwijzingen geven over regulatoire sequenties in deze genen of over de functionele domeinen van de bijbehorende eiwitten.

ABBRIVIATIONS

APO	Apolipoprotein
BP	Basepair
CETP	Cholesteryl Ester Transfer Protein
CHOL	Cholesterol
FD	Familial Dysbetalipoproteinemia
FH	Familial Hypercholesterolemia
HDL	High Density Lipoprotein
HTGL	Hepatic Triglyceride Lipase
IDL	Intermediate Density Lipoprotein
KB	Kilo Basepair
KD	Kilo Dalton
LCAT	Lecithin Cholesterol Acyltransferase
LCR	Locus Control Region
LDL	Low Density Lipoprotein
LDLR	Low Density Lipoprotein Receptor
LPL	Lipoprotein Lipase
LRP	Low Density Lipoprotein Receptor-related Protein
PCR	Polymerase Chain Reaction
QTL	Quantitative Trait Locus
RFLV	Restriction Fragnebt Length Variation
RI	Recombinant Inbred
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulphate
TG	Triglyceride
VLDL	Very Low Density Lipoprotein
VNTR	Variable Number of Tandem Repeat

CURRICULUM VITAE

Mariëtte Hoffer werd 2 juni 1965 geboren te Wieringermeer. In 1983 behaalde zij het Atheneum diploma aan de rijksscholengemeenschap Wiringherlant te Wieringerwerf. In 1983 begon zij aan haar studie biologie aan de Vrije Universiteit te Amsterdam. Het doctoraal examen medische biologie dat bestond uit de hoofdvakken Anthropogenetica (Prof. Dr. A.W. Eriksson) en Immunologie (Dr. G. Kraal en Dr. R. Scheper) werd afgelegd in 1988.

Vanaf september 1988 was zij als assistent in opleiding verbonden aan de afdeling Anthropogenetica (Prof. Dr. G.J.B. van Ommen) van de Rijksuniversiteit te Leiden. Dit proefschrift beschrijft de resultaten van onderzoek onder leiding van Dr. R.R. Frants en Dr,. Ir. L.M. Havekes (IVVO-TNO, Leiden). Sinds augustus 1993 is zij werkzaam op de afdeling Anthropogenetica van de Rijksuniversiteit te Leiden.